

Universidad Complutense de Madrid
Facultad de Estudios Estadísticos
Máster en Minería de Datos e Inteligencia de Negocios
Trabajo Fin de Máster



**Detección de omisiones y errores en datos
demográficos de cliente por su histórico de compras**

Alicia Vicente Gil
Tutora: Magdalena Ferrán Aranaz
Madrid, España
Octubre 2015

*A Rocío González, quien me enseñó y enseña
la importancia del sentido común a la hora de
enfrentarse a cualquier problema*

Índice de contenido

1. Introducción	5
1.1. Introducción a la minería de datos y los negocios	5
1.2. La necesidad del análisis de datos en el ámbito empresarial	5
1.3. Marketing, Big Data y minería de datos	7
1.4. Introducción a la empresa de estudio	8
<i>1.4.1. IKEA Portugal</i>	<i>8</i>
<i>1.4.2. IKEA Family</i>	<i>8</i>
2. Motivación	10
2.1. Situación	10
<i>2.1.1. Posicionamiento de IKEA Portugal</i>	<i>10</i>
<i>2.1.2. Valor de Cliente de IKEA Portugal</i>	<i>11</i>
2.2. Inconvenientes en la aplicación estratégica de IKEA Family	12
2.3. Objetivo	14
3. Metodología empleada	16
4. Preparación de la base de datos	19
5. Desarrollo del proyecto	24
5.1. Agrupación de los departamentos por relación de compra	24
<i>5.1.1. Análisis Factorial</i>	<i>24</i>
<i>5.1.2. Componentes principales</i>	<i>26</i>
5.2. Tratamiento de outliers	28
5.3. Agrupación de los hogares IKEA Family	30
5.3.1. Segmentación por tipología de cliente	31
<i>5.3.1.1. Conglomerados en la tipología A</i>	<i>32</i>
<i>5.3.1.2. Conglomerados en la tipología B</i>	<i>34</i>
<i>5.3.1.3. Conglomerados en la tipología C</i>	<i>35</i>
<i>5.3.1.4. Conglomerados en la tipología D</i>	<i>37</i>
5.3.2. Segmentación por conducta de cliente	39
5.4. Resultados del análisis. Clasificación de los individuos	43
6. Conclusiones y estrategias a seguir	49
7. Bibliografía	52
8. Anexos	55

Índice de figuras

Figura 1. Sistema de información de marketing

Figura 2. Banner de incentivación para captar a nuevos socios IKEA Family (web IKEA)

Figura 3. Formulario de alta como miembro IKEA Family Portugal (2015)

Figura 4. Comunicación completa IKEA (Octubre 2015 España)

Figura 5. Segmentación IKEA Family. Documento interno de estrategias

Figura 6. Stream de preparación de periodos a estudiar

Figura 7. Agregado por hogar (HHID) y cálculo de la mediana de edad por hogar

Figura 8. Stream de preparación de la base de datos a estudio incorporando la variable 'niños'

Figura 9. Nodo secuencial en el que se establece un rank decisor para deduplicar los hogares y seleccionar solo el registro del mismo hogar que primero aparezca

Figura 10. Nodo reestructurar donde creas variables procedentes de los registros de la variable DEPT_CLS

Figura 11. Stream completo de preparación de la base de datos para el análisis.

Figura 12. Gráfico de sedimentación procedente del análisis factorial

Figura 14. Descripción de componentes principales y departamentos que las forman

Figura 15. Gráficos de dispersión del gasto de los hogares en los grandes grupos. Detección de outliers.

Figura 16. -izquierda; Algoritmo bietápico utilizando solo variables porcentuales-derecha; Algoritmo bietápico utilizando solo variables de gasto

Figura 17. Centro de los conglomerados y distancia de centroides. Tipología A

Figura 18. Cantidad de gasto por departamento y clase de cliente de tipología A

Figura 19. Cantidad de hogares por grupo en la tipología A

Figura 20. Algoritmo bietápico 3 y 4 nodos para tipología A

Figura 21. Centro de los conglomerados y distancia de centroides. Tipología B

Figura 22. Cantidad de gasto por departamento y clase de cliente de la tipología

Figura 23. Algoritmo bietápico 3 y 4 nodos para tipología B

Figura 24. Cantidad de hogares por grupo en la tipología B

Figura 25. Centro de los conglomerados y distancia de centroides. Tipología C

Figura 26. Cantidad de gasto por departamento y clase de cliente de tipología C

Figura 27. Algoritmo bietápico 3 y 4 nodos para tipología C

Figura 28. Cantidad de hogares por grupo en la tipología C

Figura 29. Centro de los conglomerados y distancia de centroides. Tipología D

Figura 30. Algoritmo bietápico 3 nodos para tipología D

Figura 31. Cantidad de gasto por departamento y clase de cliente de tipología D

Figura 32. Cantidad de hogares por grupo en la tipología D

Figura 33. Cantidad de gasto por departamento y clase de cliente de todas las tipologías

Figura 34. Porcentaje de gasto por departamento de los cuatro clústeres obtenidos

Figura 35. Cantidad de hogares por tipo de cliente

Figura 36. Dendograma procedente del análisis jerárquico

Figura 37. Resultado análisis bietápico 4 nodos

Figura 38. Variables reclasificadas como nominales

Figura 39. Cantidad de hogares con hijos con o menores de 12 años en la predicción

Figura 40. Clasificación de la diferenciación de productos de niños por edad

Figura 41. Tabla de errores del árbol de decisión aplicado en ACCESORIOS

Figura 42. Tabla de errores del árbol de decisión aplicado en DORMITORIOS

Figura 43. Tabla de errores del árbol de decisión aplicado en SALÓN

Figura 44. Tabla de errores del árbol de decisión aplicado en COCINAS

Figura 45. Matriz de doble entrada del total de hogares clasificados según las predicciones y su respuesta

Figura 46. Stream de seguimiento de los hogares a lo largo de su vida que el modelo predice como no padres y ellos dicen serlo

1. Introducción

1.1. Introducción a la minería de datos y los negocios

El análisis de datos y de negocio es una disciplina que desde la popularización del *Big Data* ha experimentado una fuerte evolución en paralelo. Esto se ha debido a la convicción de que la información proporciona ventajas competitivas, lo cual provoca que aumente su importancia en el desarrollo estratégico de una empresa:

*"Dirigir bien un negocio es dirigir su futuro;
y dirigir el futuro es manejar la información" (Harper, 1961).*

La cuantía de la información disponible crece exponencialmente, con lo que el mundo está sufriendo una "*mega transformación*": desde una economía industrial a una economía basada en la información (Neisbitt, 1984). Según Nesbitt, el problema actual "*no es quedarnos sin información, sino ahogarnos en ella*". Esto es fiel reflejo de la situación actual de las empresas, siendo necesario adaptar sus máquinas para tener la posibilidad de almacenamiento y procesamiento de los datos obtenidos.

El análisis de datos ha ido evolucionando a medida que los grandes volúmenes de datos, estructurados y no estructurados crecían. La evolución ha facilitado la recolección de datos, su procesamiento y su análisis por parte de todo tipo de compañías. En este trabajo se analizará el caso de la empresa sueca IKEA, y más concretamente de su filial portuguesa. En apartados posteriores se detallará cómo llevan a cabo tanto la recolección como el análisis de los datos que van a utilizarse.

1.2. La necesidad del análisis de datos en el ámbito empresarial

A pesar de que el análisis de datos ha sido una fuente de ventajas competitivas para las empresas desde la década de los 70, solo se realizaba el análisis de los datos transaccionales, recopilados por la empresa en el transcurso de su actividad diaria. No obstante, en la última década y debido al auge de internet y las redes sociales, el *Big Data*, o tratamiento de datos a gran escala, ha pasado a formar parte de la actividad diaria de las empresas. Esto ha sido posible gracias a la reducción de costes de almacenamiento de los

datos y a la aparición de herramientas que permiten el análisis y explotación de grandes conjuntos de datos no estructurados en tiempo real.

La historia de la empresa muestra como se ha ido evolucionando en la relación empresa- cliente hasta volver de alguna manera a los orígenes, en los que el cliente vuelve a ser el centro de la actividad empresarial. Inicialmente el tamaño de las empresas era pequeño, como las clásicas tiendas de barrio donde el tendero conoce a todos sus clientes, junto a sus gustos, situaciones familiares y preferencias. Poco a poco, este concepto de negocio evolucionó y surgieron las grandes superficies comerciales, donde la relación tendero-cliente se despersonalizó y se pasó a comprar a un ente y no al comerciante de toda la vida. Esto generó grandes beneficios, ya que estos negocios se podían permitir competir en precios porque se beneficiaban de economías de escala, mientras que los pequeños comercios no podían. Esta es la causa de que progresivamente estas tiendas hayan sido desbancadas de sus puestos de suministradores.

No obstante, las grandes empresas se encontraron con que para poder competir con sus iguales, necesitaban recuperar el poder que tenían los pequeños comercios: conocimiento detallado de la información de cliente, que permite la personalización de los servicios. Esto comenzó a ser una realidad tras la aparición de empresas como *Business Objects Actuate*, *Crystal Reports* o *Micro Strategy*, entre otras muchas, que ofrecían a las grandes empresas informes y análisis de sus datos, que les permitían recuperar la cercanía al cliente y el conocimiento de sus necesidades y preferencias.

Esta evolución ha motivado a muchas firmas a la creación de planes de fidelización del cliente con los que conseguir sus datos demográficos y de compras para poder proceder a su análisis y aumentar su ventaja competitiva derivada de la posesión de información útil. Así pues, hoy en día se observa la lucha de las diferentes empresas de *retail* por conseguir un hueco en los tarjeteros de sus clientes.

La búsqueda de recolección de datos tiene entonces lugar como medio para que las empresas tengan la posibilidad de acercarse a sus clientes (todo lo que permiten las leyes), y así, poder ofrecerles, a través de distintos impactos y según las características de estos, ofertas personalizadas, ya no solo por características demográficas, sino por el comportamiento que se les predice a través del histórico de compras que las tiendas almacenan. Esta cantidad ingente de datos, a veces inmanejable, permite, a través de un

fuerte tratamiento de normalización, y conocimiento del negocio (cualquier minero de datos que se considere bueno tiene la necesidad de conocer y entender sus datos, sino, no tendrá sentido su trabajo) la personalización de los impactos, consiguiendo influenciar al cliente de la forma buscada.

1.3. Marketing, Big Data y Minería de datos

Según Kotler (2003), *"un sistema de información de marketing (SIM) está formado por un conjunto de personas, equipos y procedimientos capaces de recoger, clasificar, analizar, evaluar y distribuir la información oportuna y precisa para la toma de decisiones"*. Cuanto más complejas son las actividades de marketing utilizadas por la empresa y mayor es la competitividad de ésta, se necesita una mayor cantidad y calidad de datos para obtener como output una información útil que se traduzca en resultados de negocio. No basta con poseer una gran base de datos, sino que es necesario un buen equipo que sepa tratar, analizar y explotar esa información para conseguir explotar el potencial de la base de datos.

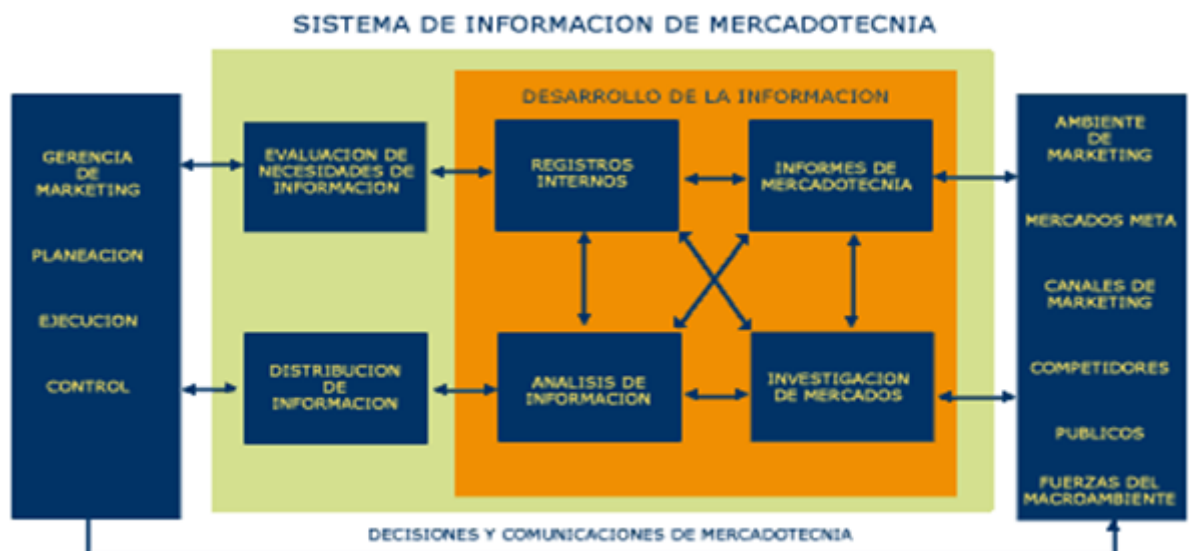


Figura 1. Sistema de información de marketing.

1.4. Introducción a la empresa de estudio

IKEA es una multinacional de origen sueco, líder en el mercado de la distribución de mobiliario y objetos de decoración para el hogar. Oferta a sus clientes una gran gama de productos funcionales, de calidad, diseño nórdico y a precios asequibles.

El Grupo IKEA, propiedad de la Fundación Stichting INGKA, tiene internalizada toda la cadena de valor, desde la estrategia del surtido y el desarrollo de productos, hasta la fabricación, distribución y venta. El Grupo IKEA tiene un total de 288 tiendas en 27 países y 35 más operando en régimen de franquicia.

1.4.1. IKEA Portugal

En Portugal, IKEA dispone de 3 tiendas. La primera, ubicada en Alfragide, abrió sus puertas en Junio de 2004. Tres años después, se inauguró una nueva tienda en Matosinhos y en 2010 se abrió la tercera tienda en Loures. Además, IKEA Portugal continuará con su expansión geográfica, inaugurando nuevas tiendas en Braga y Leule. Estas aperturas tan cercanas en el tiempo han sido posibles gracias a los beneficios obtenidos por la marca en el país vecino, debidos a la gran acogida de IKEA por parte de los clientes portugueses. El éxito en este país, ha podido analizarse gracias a los datos aportados por el plan de fidelización de IKEA.

1.4.2. IKEA Family

IKEA se encontró con un problema clásico de una multinacional: los conocimientos que tenía de sus clientes eran meramente intuitivos y se obtenían a través de la generalización de resultados de estudios de mercado muy básicos. Es por esto, que en línea con la tendencia generalizada de incrementar la información útil de los clientes por parte de las empresas, descrita en párrafos anteriores, que surge el plan IKEA Family.

IKEA Family es un club de fidelización que da ventajas a los miembros por presentar su tarjeta en cada compra en IKEA. La empresa invierte parte de sus ingresos en

el Plan de Fidelización IKEA Family, con el fin de conocer las características de los individuos a quienes se dirige, de lo que se beneficia la empresa en su conjunto. Además, con una fidelización efectiva se produce un aumento de clientela, de gasto medio por cliente, de identificación del cliente con la marca, una mejora del posicionamiento, etc.



Figura 2. Banner de incentivación a para captar a nuevos socios IKEA Family (web IKEA).

La tarjeta de socio se obtiene tras rellenar los datos de usuario en máquinas dispuestas en la tienda, en la web o rellenando el formulario en papel, y gracias a ella, los miembros IKEA Family pueden beneficiarse de muchas ventajas exclusivas. En Portugal se ha llegado al millón de socios en mayo de 2015 y actualmente esta cifra sigue subiendo. El formulario que se debe rellenar para ser miembro IKEA Family es el siguiente:

1 Preencha o formulário de adesão online.

2 Imprima o seu cartão num quiosque IKEA FAMILY disponível na sua loja IKEA.

3 Apresente o seu cartão IKEA FAMILY sempre que nos visitar.

Formulário de Adesão

Os campos com * são de preenchimento obrigatório para a ativação do seu cartão IKEA FAMILY. Consulte as [Condições Gerais](#)

Nome completo *

Nome para o cartão *

Data de nascimento *

E-mail *

Confirme o seu e-mail *

Telefone móvel *

Morada completa *

Tem 2ª habitação?

Situação familiar

Nº de filhos

Sexo ☐ M ☐ F

NIF *

Telefone fixo

Nome da rua

Número **Andar** **Localidade**

C.P. **C.P.** **Localidade**

Tem negócio próprio? ☐ Sim ☐ Não

Casado(a) / União de facto ☐ **Sozinho(a)** ☐ **Partilho casa com amigos / familiares** ☐

Área da sua casa (m2) **Há quantos anos vive na sua casa?**

De forma a finalizar a sua adesão, por favor crie uma password.
Esta password, juntamente com o seu e-mail será a chave de acesso à sua área pessoal IKEA FAMILY. A password deverá ter entre 6 e 10 caracteres.

Crie uma password * **Confirme a password ***

Palavra de Controlo

Digitar o texto da imagem à direita garante que uma pessoa, e não um programa automático, está a fazer esta adesão.

[alterar palavra](#)

☒ Autorizo a IKEA PORTUGAL - MÓVEIS E DECORAÇÃO, LDA. a contactar-me, designadamente, através de meios de comunicações eletrónicas para me informar de ofertas promocionais e outras ações de marketing.

☐ Li e aceito as [Condições Gerais](#).

[cancelar](#) [aderir](#)

Figura 3. Formulário de alta como membro IKEA Family Portugal (2015)

2. Motivación

2.1. Situación

2.1.1. Posicionamiento de IKEA Portugal

En la sociedad *sobre-comunicada* en la que vivimos, todas las marcas buscan conseguir impactar al cliente de forma que sobrepasen las barreras comunicativas que ellos mismos crean para sobrevivir a los impactos constantes que reciben. Este es el motivo, por el que IKEA busca conseguir evitar estos muros a través de un buen posicionamiento de marca. El posicionamiento de IKEA tiene su origen en Ingvar Kamprad, su fundador, con su primera tienda en Småland, Suecia, en 1943.

Como mejor forma de acercarse a sus consumidores, IKEA quiere llegar a través de un único y acertado impacto, que se consigue con un posicionamiento de marca que esté unido a un *atributo esencial*. En el caso de IKEA, se ha elegido el posicionamiento de precio bajo cuidando la calidad. IKEA se caracteriza por su mítico mensaje de **“precios bajos con significado”** y la marca se preocupa por ofrecer el concepto *de low price with a meaning*, ya que busca cumplir su lema *"IKEA is for the many"*, ofreciendo productos de diseño, accesibles para todos los bolsillos y que respondan a las diferentes necesidades de decoración que se van desarrollando a lo largo de la vida del cliente.

El posicionamiento es un proceso subjetivo que tiene lugar en la mente de los consumidores, e IKEA lo sabe. Es por ello que en todas sus campañas busca que sus clientes sean conscientes de que la marca es familiar, algo que cualquier familia de cualquier estilo se puede permitir. IKEA vende un hogar donde sentirse a gusto, desenfadado, y así lo comercializa en sus creatividades: *"un hogar, no solo muebles"*, y esto es lo que se pretende hacer llegar a la mente de los clientes cuando piensan en IKEA, *"un hogar a un precio aceptable"*.



Figura 4. Comunicación IKEA (Octubre 2015 España)

Además de un mensaje bien definido, la empresa sueca se esfuerza en mantener el posicionamiento que le ha llevado muchos años conseguir gracias a su fuerte notoriedad, ya que la empresa es conocida mundialmente.

IKEA llegó a Portugal en 2004 y, a pesar de no ser una de las empresas de muebles tradicionales y con más trayectoria del país, gracias a su fuerte inversión en comunicación ha conseguido que la mayoría de portugueses hayan sucumbido al estilo y concepto de decoración que los países escandinavos ofrecen. De hecho, al preguntar por una empresa de muebles, será una de las primeras en venir a la cabeza de cualquier individuo, sea *pro IKEA* o no, ya que la marca es un *top of mind* indiscutible. Esta información la ratifican datos como los obtenidos por la empresa Pentasoft, una de las empresas encargadas de ejecutar las estrategias de fidelización del club IKEA Family que realiza IKEA Portugal.

2.1.2. Valor de Cliente de IKEA Portugal

IKEA busca el contacto más efectivo con sus clientes. La empresa utiliza distintos métodos para contactar con los individuos:

- Correo ordinario: *Now* 's y la revista de IKEA
- SMS
- Email.

IKEA no solo selecciona el canal por mayor efecto en el cliente, también lo selecciona por el valor que tenga cada hogar para la marca. Esto se debe a que los distintos canales de comunicación llevan asociados distintos costes de facturación. La revista y el *now* son los canales más caros, siendo el email el más económico. Por ello, hay que evaluar que clientes merecen una mayor inversión en su contacto. La empresa sueca no solo debe evaluar individualmente el valor de cada cliente, sino que la valía del análisis reside en la comparativa entre ellos, ya que las cantidades de envíos están limitadas a un determinado número, por lo que se necesita un ranking que indique que clientes son mejores.

Además, hay impactos que solo deben recibir determinados grupos, dado el objetivo que se persiga con ellos: reactivar, aumentar el gasto, potenciar la venta cruzada, etc. Por ello, Customer Value, la empresa dedicada al análisis, procesamiento y estudio de los datos de IKEA Family, ha realizado una segmentación con la que clasificar a los individuos según su potencial, y dentro de ella, un ranking para diferenciar a los mejores de cada grupo.

Así, se pueden diferenciar cuatro grupos de clientes;

- **Clientes de la tipología A:** Son individuos con un gasto medio alto, con compras en las familias importantes de muebles y compradores de muebles.
- **Clientes de la tipología B:** Son individuos con un gasto medio normal, compradores de muebles (en su mayoría) y accesorios.
- **Clientes de la tipología C:** Son individuos con un gasto medio bajo y compradores de pequeños muebles y accesorios.
- **Clientes de la tipología D:** Son individuos con un gasto medio bajo y compradores solo de accesorios.

Estos grupos serán utilizados en el desarrollo del proyecto que trata este trabajo de fin de máster.

2.2. Inconvenientes en la aplicación estratégica de IKEA Family

Puesto que la filial es consciente de su posicionamiento, busca asegurarlo a través de una relación más directa con sus clientes, diferenciando sus impactos según qué tipo de familia tenga cada hogar, es decir, personalizando su comunicación para conseguir crear en el consumidor una sensación de mayor cercanía con la marca, y con ello, aumentar sus beneficios. Como se ha descrito en apartados anteriores, esto es posible gracias al club de fidelización IKEA Family, que permite a la empresa tener toda la información que el consumidor y miembro del grupo está dispuesto a dar, y utilizarla para un estudio más exhaustivo, a través del cual, la empresa pueda conocer mejor a sus clientes y poder así

adaptarse a sus necesidades. IKEA Family divide a sus clientes en los siguientes tipos de familia o situaciones de vida:

- *Living with children*
 - *Baby*
 - *Toddler*
 - *Starting school*
 - *Tweens & teens*
- *Living without children*
 - *Living single*
 - *Living single starting out*
 - *Living single stabilished*
 - *Living together*
 - *Living together starting out*
 - *Living together, stabilished*



Figura 5. Segmentación IKEA Family. Documento interno de estrategias.

Con el formulario de alta como miembro Family se consiguen todos los datos de contacto que interesan a IKEA: dirección de correo electrónico, dirección postal y número de teléfono. Además, también se consiguen datos demográficos, lo que permite dividir a la población según sus características demográficas: edad, sexo, situación familiar, número de hijos, negocio, tamaño del hogar y años en su hogar.

Suponiendo que los individuos dicen la verdad, se pueden realizar muchas segmentaciones de los miembros IKEA Family a la hora de ofrecerles determinados productos y ofertas, teniendo una mayor probabilidad de acierto y de conseguir que estos clientes acudan a la tienda y compren. Sin embargo, los datos demográficos no bastan, e IKEA necesitaba comprobar si perfiles demográficos similares indicaban comportamientos de compra iguales. Se descubrió que esto solo ocurre en cierta medida y en determinados casos. Por ello, a lo largo de los años y a través de un proceso de prueba y error, ha ido cambiando la cantidad y el tipo de datos que se piden en el test de alta.

Uno de los problemas a los que se enfrenta IKEA Family es la actualización de los datos. Es decir, no se puede conocer cómo evoluciona la situación de cada miembro a lo largo de su ciclo de vida como cliente si no se le indica. Esto provoca que haya una

distorsión de la realidad entre la información que extrae IKEA de sus datos y la situación real del cliente. Este problema no puede evitarse y siempre va a existir este tipo de error, sin embargo, se intenta minimizar a través de recordatorios y concursos para que los clientes actualicen sus datos.

2.3. Objetivo

Las *living situations* o situaciones de vida, como se muestra en el apartado anterior, son una clasificación basada en variables socio demográficas que IKEA utiliza para acercarse y conocer más a sus clientes. De hecho, las tiendas están decoradas respondiendo a estas situaciones de vida.

Si se observa dicha clasificación, la primera división crucial se hace en base a que en la familia haya o no niños, es decir, la primera división que marcará grandes diferencias entre un grupo y otro es el concepto de *living with children*. Por este motivo, es importante conocer si en el hogar hay niños o no, y se sabe que esta información no siempre está recogida de manera fidedigna en la base de datos.

A consecuencia de ello, con este estudio se pretende abordar el problema de error en los datos obtenidos por la pregunta de *número de hijos* y dar un paso más orientado a subsanar las deficiencias lógicas que se presentan.

Con esta pregunta, se van a encontrar distintos tipos de error:

- **Error por no actualizar los datos;** Los miembros de IKEA Family se dan de alta con una determinada situación de vida: no tienen hijos cuando se dan de alta y años después pasan a tenerlos, pero IKEA lo desconoce ya que no han actualizado esta información en su perfil.
- **Error por omisión de información;** la respuesta a esta pregunta no es obligatoria para hacerse miembro de IKEA Family, por lo que un nuevo miembro puede decidir no dar este dato.
- **Error por mentir;** Un nuevo miembro puede decidir mentir en sus respuestas del cuestionario.

Para IKEA es muy importante conocer qué miembros tienen hijos y cuáles no, ya que muchas de las campañas que se hacen se dirigen específicamente a los hogares con niños. No solo eso, este campo influye en el primer paso de la personalización de los impactos de la filial portuguesa hacia sus clientes, ya que las creatividades se personalizan en función de que haya o no niños en el hogar. Los impactos se diferenciarán a través de fotos de familias según tengan o no hijos, productos ofertados dirigidos a familias con o sin niños, recomendación de talleres infantiles o para adultos...

Al analizar los casos de empresas como *Amazon* y su sistema de recomendación, *Coca Cola* o *Nutella* con su personalización de envases, entre otras, queda claro que la personalización de los impactos, sea de una manera (producto directo) u otra (envíos publicitarios), es una buena estrategia. Sin embargo, si antes no se reclasifica a los individuos erróneamente ubicados, la diferenciación será un esfuerzo, tanto económico como humano, que no tendrá la recompensa y resultados que podría tener si se estabilizan los datos. En consecuencia, se obtendrá un *ROI (Retorno de la Inversión)* más bajo de lo esperado y no se optimizará la experiencia cliente.

Así, lo que se busca es conseguir una base de datos más pura para proceder a desarrollar los planes de marketing personalizados y con ello, aumentar los beneficios que produce IKEA Family. Esto se conseguirá creando un árbol de decisión que perfile a los clientes en con o sin niños por su comportamiento de compra y no solo por sus características demográficas. Con ello, se obtendrán cuatro grupos de hogares según su respuesta y clasificación del modelo:

- **Padres reales**, que dicen tener hijos y el modelo los clasifica como con niños.
- **Potenciales padres**, que dicen no tener hijos pero el modelo los clasifica como con niños.
- **No padres**, que dicen no tener hijos y el modelo los clasifica como sin niños.
- **Padres con potencial de compra**, que dicen tener hijos pero el modelo los clasifica como sin niños.

Gracias a esta diferenciación, se conseguirá el objetivo de estudio, la reclasificación de los hogares según tengan o no niños en el hogar.

3. Metodología empleada.

En este apartado se muestra el camino a seguir para la realización de este proyecto.

Se comienza con la preparación de los datos, estos, se encuentran en tres archivos distintos.

- *PT_Clientes*
- *PT_Tipologia*
- *PT_Ventas_totales*

Todos ellos tienen variables de unión como lo es el identificador de hogar o el *cardnumber*. A través de *SPSS Modeler* se realizarán uniones de registros, creación de nuevas variables, y transformación y eliminación de otras, para conseguir separar en un solo archivo las variables que interesan, así como periodos de actuación, clientes y uniones de estos. Una vez el fichero esté listo, se exportará a un fichero de tipo *SPSS Statistics* para poder proseguir con el trabajo. La utilización de las dos herramientas se debe al mayor potencial de *Modeler* para tratar las bases de datos y al de *Statistics* para modelar los datos.

Una vez creado el archivo, con *SPSS Statistics* se aplicarán componentes principales para evaluar las distintas variables y seleccionar las que interesa mantener y agruparlas para simplificar el estudio. Si los resultados son interpretables y coherentes a nivel de negocio, con o sin la aplicación de rotación de los ejes, se mantendrán estas componentes, con lo que se conseguirá una reducción de variables que permita el enfrentamiento de éstas dos a dos para la búsqueda de los registros *outliers*, y a través de la aplicación de filtros en *SPSS Modeler* sobre la cantidad máxima de gasto de los hogares en los departamentos que forman cada componente, se limpiará el ruido que provocan en el análisis.

Con una base de datos en la que los hogares tienen comportamientos *normales*, es posible aplicar algoritmos de conglomerados que permitan agrupar a los individuos en grupos donde todos los registros pertenecientes a ese segmento sean similares entre ellos y diferentes a los registros pertenecientes a otros clústeres.

Gracias a la limpieza de *outliers* se consiguen evitar los *segmentos satélite* o, por lo menos, disminuirlos, lo que simplificará mucho el trabajo.

Es imprescindible tener en cuenta que los clientes son muy distintos, tanto en porcentajes de gasto por departamento, como en cantidad de gasto, debido a distintos tamaños de la cartera de los clientes y de su unión con la marca. Es por esto que se procede a través de *SPSS Modeler* a dividir a los clientes en los 4 grupos establecidos en IKEA Family de los que se ha hablado anteriormente (tipología A, B, C y D).

La división se realiza para utilizar las variables porcentuales de gasto, en vez de las variables de cantidad de gasto, así como de los dos tipos a la vez, ya que lo que se busca es un perfil de consumo (proporción) y no de gasto, ya que un hogar con o sin niños puede tener cualquier nivel económico. Con esto se consigue una diferenciación manual de calidad de cliente, y se procede a una clasificación programada de porcentaje de gasto.

Una vez diferenciados los cuatro grupos, se procede con *SPSS Statistics* a aplicar algoritmos no jerárquicos y bietápicos dentro de los cuatro grupos. El bietápico se utiliza como indicador de calidad de las distintas elecciones de cantidad de clústeres finales. Esto se hace porque los algoritmos no jerárquicos necesitan que se seleccione la cantidad de clústeres final deseada. Se escogerán las cantidades de clústeres que mejores indicadores den y que tengan sentido de negocio.

La solución óptima sería utilizar el algoritmo jerárquico en los 4 grupos (tipologías de cliente) en vez de no jerárquicos, sin embargo, el algoritmo jerárquico presenta un problema: con una base de datos de más de 700.000 registros no funciona apropiadamente con los programas en los que se ha probado (*SAS Miner*, *SAS Base*, *SPSS Statistics* y *SPSS Modeler*). Es por ello que se utiliza en este paso, una vez reducidos los datos con los que trabajar, aplicándolo a una matriz manejable compuesta por los diferentes clústeres y las variables procedentes de las componentes principales. Con la aplicación de este algoritmo se conseguirá la unión de los clústeres más próximos entre sí y, con ello, la segmentación final de los individuos del universo de estudio.

Tras la obtención de la segmentación final, se consigue el perfil de compra de los distintos grupos de clientes, tengan una calidad u otra, ya que no hay que olvidar que el objetivo de este análisis es conocer el dato demográfico de niños, no conocer la calidad del cliente.

Una vez divididos los clientes por similitud de gasto y porcentaje de gasto, se realizarán las medias de consumo por departamento y se crearán nuevas variables que indiquen si un hogar es o no cliente de un departamento. Éstas, junto a la *edad*, son las variables que se introducirán como independientes en el árbol CHAID-Exhaustivo, y como dependiente la variable *hijos* con objetivo de respuesta.

Gracias a los árboles de decisión, se consigue clasificar a los registros en grupos distintos hasta identificar las respuestas de la variable objetivo. En nuestro caso, se obtendrán cuatro árboles distintos que clasificarán a los individuos en los cuatro tipos de familia buscados, gracias a sus predicciones de '*si*' y '*no*'.

4. Preparación de la base de datos

Se tienen los ficheros de *PT_Clientes*, *PT_Ventas_totales* y *PT_Tipología*, donde se tienen las variables necesarias para el estudio (anexo 1).

IKEA sabe que el cliente se dirige a la tienda en sus cambios de vida. Es por eso que, según estudios anteriores, se conoce que el primer año en el ciclo de vida de un miembro IKEA Family es el más activo y en el que más compras realiza. Por este motivo, se analizará solo el primer año de vida de los clientes. Además de este criterio de actividad, esta decisión se toma por la necesidad de simplificar el estudio, ya que es una cuantía de datos que las máquinas que actualmente se tienen a disposición no permiten su modelado.

A continuación se procede a la separación del primer año de vida de cada hogar. Debe tenerse en cuenta que en IKEA se trabaja a nivel hogar y no a nivel individuo, debido a que IKEA vende muebles para el hogar, por lo que dos impactos en individuos del mismo hogar únicamente aumentaría los gastos de marketing sin atraer a más clientes. Esto significa que, pese a que todos los miembros de un hogar pueden ser clientes Family y tener compras en sus distintas tarjetas (un individuo puede tener varias tarjetas con distinto número), IKEA los une a nivel hogar (por dirección, teléfono y apellidos), y las identifica a través de la variable hogar, la *HHID*.

Para realizar esta selección del periodo de estudio, hay dos puntos importantes que tener presentes:

- Los individuos del hogar pueden tener fechas de alta o de primera operación distintas, por lo que no es posible realizar un corte horizontal en el tiempo. Hay que calcular la fecha mínima entre todas las tarjetas del hogar y por hogar.
- Estudios internos de IKEA han mostrado que una buena cantidad de clientes se dan de alta en IKEA Family al salir de la tienda, por lo que la fecha de activación y de primera operación no es la misma. Esto penaliza a estos individuos en cuantía de compras del primer año como miembros Family. Por esto, se selecciona la fecha mínima de la primera operación y no se tiene en cuenta la fecha mínima de alta del hogar.

A través del fichero PT_Ventas, se obtiene una primera rama con un nodo agregado en el que se agrega solo por hogar, y se incluye el cálculo del TRX_DT_Min (fecha mínima de operación). Se vuelve a fusionar por hogares coincidentes con el fichero de ventas inicial. Así, lo que se consigue es mantener el fichero completo inicial pero agregando una variable más (TRX_DT_Min). Se tiene la problemática de que la fecha de compra es de tipo *fecha*, por lo que se desagrega en día, mes y año, y se convierten todas ellas (3 variables nuevas) en tipo *cadena* con la utilización de *nodos derive*, que permite generar nuevos campos a través de otros ya existentes.

1. Nodo *derive* cambio de la fecha a año y paso a tipo *cadena*.
2. Nodo *derive* cambio de la fecha a mes y paso a tipo *cadena*.
3. Nodo *derive* cambio de la fecha a día y paso a tipo *cadena*.
4. Nodo *derive* unión de día mes y año+1 a una fecha.

Una vez se tienen estas cuatro variables nuevas, se anulan con un filtro la de año, mes y día y se realiza una selección del periodo deseado (primer año de vida como cliente del hogar) con un nodo *generar* donde se introduce la *query* de $TRX_DT \geq TRX_DT_Min$ and $TRX_DT \leq 'fecha+1'$ y se impone la opción de *incluir estos casos*.



Figura 6. Stream de preparación de periodos a estudiar.

Con esto, el siguiente resultado se queda solo con las compras de cada hogar en su primer año de vida como miembro de IKEA Family.

Una vez determinado el periodo, hay que comenzar a pensar en las variables que se podrán necesitar en el estudio de predicción de *padres* por histórico de compra. Se necesitarán las variables provenientes de *PT_Ventas*, sin embargo, para su posible futura interpretación, será necesario simplificar. Es decir, no utilizar a nivel referencia de producto ya que si fuera así, la interpretación sería inviable. Esto también sucederá si se

realiza por nombre o familia de producto: no hay que olvidar que IKEA se caracteriza por sus especiales nombres para sus productos.

En consecuencia, se selecciona la variable *DEPT_CLS* (número de departamento) y se realiza un agregado por hogar, *dept_cls* y *purch_amount sum* (identificador, departamento y gasto total de cada hogar en cada departamento). Así, se consigue una línea por hogar y departamento comprado. Es decir, si un hogar ha comprado en tres departamentos distintos, tendrá tres líneas, cada una con su número de departamento y gasto total de ese hogar en ese departamento. Actualmente se tienen las variables:

- Hogares (identificadores).
- Departamentos de compra.
- Cantidad en euros gastada.

Se plantea qué datos pueden ser interesantes para el análisis. Se llega a la conclusión de que la edad puede ser una variable interesante y que saque de dudas a la hora de establecer si un hogar tiene o no niños. Esto es totalmente relativo, puesto que cada día, el abanico de edad en el que se tienen los hijos es más grande. Sin embargo, la variable se introduce al análisis. Ahora bien, se debe establecer una edad por hogar para no duplicar resultados. De esta forma, nos encontramos ante dos posibilidades: realizar la mediana de la edad del hogar o utilizar la edad máxima estableciendo prioridad a la edad de la mujer, ya que se conoce que la mujer es la que es más propensa a comprar en IKEA, o por lo menos a utilizar la tarjeta Family. En este caso utilizaremos la mediana de la edad del hogar, ya que se ha considerado que son los propietarios y centros del hogar (núcleo paterno) los que tienden a comprar en IKEA y no las otras ramas de la familia (hijos, tíos o abuelos).

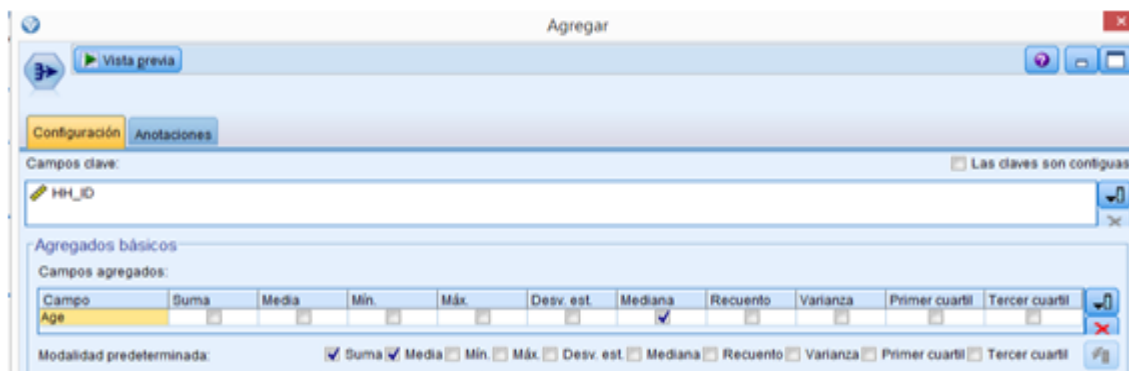


Figura 7. Agregado por hogar (HHID) y cálculo de la mediana de edad por hogar.

Otro factor interesante es la variable tipología, de la que ya se ha hablado con anterioridad. Con ella se encuentra un problema, y es que no se puede utilizar la tipología calculada con datos totales de venta, puesto que la tipología es variable. Es por eso que se procede al cálculo de la clasificación de los individuos a través de la tipología basada en el archivo obtenido de vida de clientes en el primer año. Como cada miembro de un mismo hogar, debe de tener la misma tipología asignada, no se obtienen duplicados.

Una vez se tiene la *edad* y la *tipología*, se añade la variable *niños*, que es la variable que se busca predecir. La variable *with children* (procedente del archivo *PT_Customers*), no tiene porqué ser igual en dos individuos del mismo hogar. Sin embargo, se asume que cuando alguien se toma el tiempo suficiente para rellenar el nombre y fecha de nacimiento de un hijo, es un suceso cierto, con mayor certeza que quien no contesta. La no contestación puede significar que no tiene hijos, que miente o que se niega a responder.



Figuro 8. Stream de preparación de la base de datos a estudio incorporando la variable 'niños'.

Por este motivo se realiza un agregado por hogar y *with children* (cuarto nodo de la imagen superior) del fichero *PT_Cientes*. Se añade a este un nodo *ordenar* por *hhid* ascendente y por *with chidren* descendente (quinto nodo). Esto significará que las repeticiones de hogar saldrán juntas, y en primer lugar el individuo del hogar que tiene un *sí* en la variable *with children*.

Una vez se tiene el archivo en este orden, se procede a realizar un *rank decisor* (sexto nodo). Es decir, la creación de una variable que pone uno al primer valor de cada hogar y cero al resto. Seguidamente, se añade un filtro en el que se genera la *query* de "*rank decisor* = 1" (séptimo nodo).

Así, se obtiene un archivo con la variable

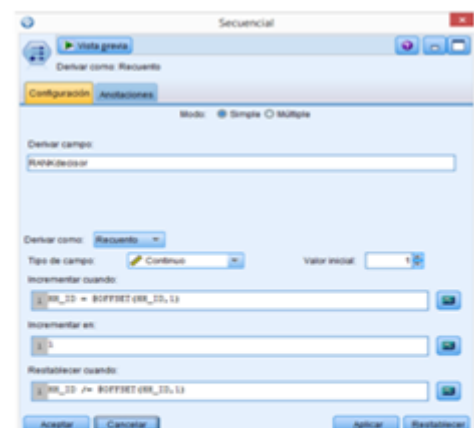


Figura 9. Salida de nodo secuencial en la que se establece un Rankdecisor para poder deduplicar los hogares y seleccionar solo el registro del mismo hogar que primero aparezca.

hhid, *niños* y *rank decisor*, donde solo puede haber una línea por hogar, todo el *rank decisor* tiene que repetir el valor uno y si un hogar tenía un *sí* y un *no*, solo se ha quedado con el *sí*. Una vez se ha construido esta variable, se elimina la variable *rank decisor* y se une al fichero *edad* a través de la fusión por *hhid* coincidentes con este.

Con este paso, se ha conseguido un archivo con *identificador (hhid)*, *niños*, *edad*, *tipología*, *gasto total por departamento y por hogar y el identificador del departamento*.

Sigue existiendo la dificultad de duplicación de filas del mismo hogar debido a

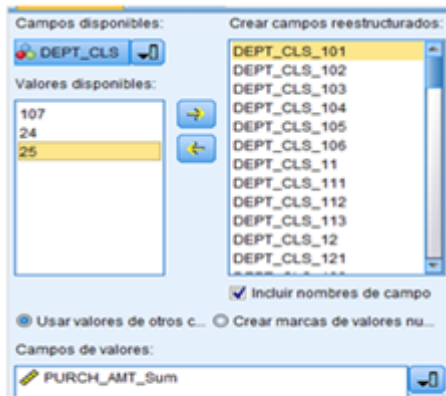


Figura 10. Salida del nodo reestructurar donde creas variables procedentes de los registros de la variable *DEPT_CLS*

que son clientes de distintos departamentos. Por este motivo, se procede a realizar una transposición de los datos de la matriz. Es decir, se van a convertir en variables lo que actualmente son identificadores en la variable departamento. Así pues, se obtiene una matriz con tantas filas como hogares con compras hay, y con una cantidad de variables de *hhid*, *tipología*, *edad*, *niños* y *gasto total por deptX* (todos los departamentos a utilizar). Esto se realiza con el nodo *reestructurar*.

Este cambio provocará blancos o nulos en las variables de gasto en los departamentos donde un hogar no haya comprado. Por ello, se añade el nodo *reclasificar* y se indica que todo registro con blancos o nulos en alguna de las variables de cantidad o gasto, se convierta en cero.

Se incorpora también la variable *GASTO TOTAL*, que es la suma de todos los gastos totales de los departamentos para posibles usos. A continuación se guarda la tabla obtenida con el siguiente *stream*, la cual será la base del proyecto.



Figura 11. Stream completo de preparación de la base de datos para el análisis.

5. Desarrollo del proyecto

Obtenida la base de datos de 82 variables y 766.116 registros distintos (hogares), se procede a su tratamiento y explotación como se mostrará en los siguientes puntos.

5.1. Agrupación de los departamentos por relación de compra

Hay tres tipos de variables en el archivo, la variable identificadora, las demográficas y las de compra. Las variables demográficas se utilizarán más adelante cuando se busque realizar clústeres de los hogares clientes. En este punto, nos centramos en las variables procedentes del histórico de compras.

82 variables no es una cantidad manejable, por lo que se comienza con la disminución de este número a través de agrupaciones en variables representativas de éstas. Esto se hace aplicando el *análisis factorial* o el *método de componentes principales*.

5.1.1. Análisis Factorial

Tanto el análisis factorial como el análisis de componentes principales son métodos multivariantes de reducción de dimensión del problema. Es decir, estos métodos combinan distintas variables para obtener una cantidad inferior de variables ficticias que representen a las variables que las forman perdiendo la menor cantidad de información posible. El análisis de componentes principales se utiliza cuando la varianza de cada variable original se explica totalmente por las variables cuya combinación lineal determina sus componentes. Esto no ocurre en el análisis factorial. Es por esto que probaremos el análisis factorial antes.

Este método se aplica para reducir la cantidad de variables que son de datos cuantitativos correlacionados entre sí, y agruparlas en unas pocas mayores denominadas factores, que sinteticen la mayor parte de la información de las variables que las forman. Es decir, que tenga sentido y un alto volumen de representatividad.

En este paso, solo se utilizarán las variables numéricas por dos motivos.

1. No es posible utilizar nominales en componentes principales ni en factorial.
2. No es su fin, ya que lo que se busca en este paso es reducir el tipo de cliente por departamento. Por ejemplo, "cliente de dormitorio", en vez de tener el cliente de colchones, el de camas, el de armarios, el de cómodas, el de iluminación o el de escritorio por separado.

A continuación se realiza el estudio con las variables de gasto por departamento. El resultado obtenido muestra un valor alto del KMO con 0.9, un determinante de la matriz de correlaciones muy pequeño (8.098×10^{-7}), y un p valor igual a 0, lo cual muestra que los datos son adecuados para realizar un análisis factorial.

La varianza acumulada es muy pequeña, 31.68% e indica que lo óptimo son 8 factores puesto que incrementos en el número de factores apenas aumentan la varianza.

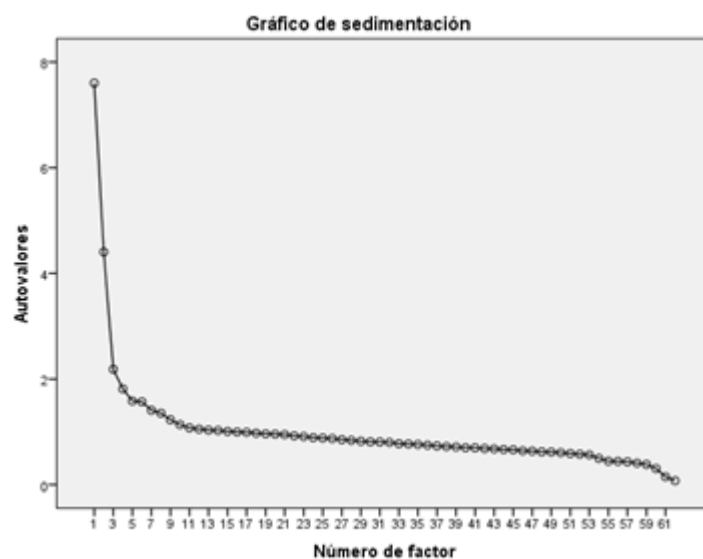


Figura 12. Gráfico de sedimentación procedente del análisis factorial

Esto se observa en el gráfico de sedimentación, que indica que solo 8 factores tienen autovalores superiores a 1. El valor del porcentaje explicado de la variabilidad es muy bajo, por lo que se procederá a la eliminación de variables con coeficientes KMO bajos en la matriz anti imagen . Se observan distintas variables con valores inferiores a 0.5 en esta matriz, por lo tanto, eliminables. Se ha procedido a eliminarlas una a una del estudio hasta quedarnos con la solución óptima. Se han eliminado del archivo original las

variables *PURCH_AMOUNT_SUM* y los departamentos; 62, 72, 74, 111, 131, 71, 73, 77 y 79.

Con esto se consigue un KMO= 0.91, un determinante de 0.001, una significación de Bartlett = 0, lo que hace que el análisis factorial sea válido.

Se continúa con el estudio y se observa la variabilidad explicada acumulada, solo se consigue un 25% con variables con autovalor superior a 1. Esto indica una mala calidad de las futuras predicciones y de la representatividad de estos factores. Así pues, se para este análisis y se procede a aplicar componentes principales.

5.1.2. Componentes principales

Se trabaja con las variables procedentes del gasto de cada departamento, ya que se busca conseguir componentes principales de las variables de gasto de clientes en cada uno de ellos. Como no son comparables, ya que el gasto de comprar un mueble es mucho más alto que el de comprar accesorios, se utiliza la matriz de correlación en vez de la de covarianzas.

La teoría indica que solo se deben retener las componentes que explican la mayor parte de variabilidad de las variables iniciales. Se realizaron muchas pruebas en las que se eliminaban variables que tenían un coeficiente KMO muy bajo para comprobar si las componentes mejoraban. Sin embargo, esto no era necesario, ya que el coeficiente KMO con todas las variables superaba el límite que establece la teoría de 0.6.

Las pruebas indican que los resultados no mejoran por eliminar variables, y no solo eso, desde el punto de vista de negocio, no es lógico eliminar varios departamentos pese a que el KMO lo indique en la matriz anti imagen. Los resultados detallados del método de componentes principales seleccionado se encuentran en el anexo 2.

El KMO es de 0.9, la significación tiene un valor de 0 y se consigue explicar con 15 componentes una variabilidad del 47%, lo cual no es muy alto pero comparado con el resto de resultados anteriores, es la elección óptima. Así pues, se prosigue con el análisis y se comprueba que la matriz de componentes principales sin rotar no tiene demasiado

sentido, es por esto, que se establece una rotación con el método *varimax*. Las rotaciones de los ejes son un método muy utilizado para buscar interpretabilidad de las componentes, además, no cambian la proporción de inercia explicada de las variables por parte de las componentes.

Se aplica la rotación de tipo *varimax*, ya que se ha considerado que es la que mejores resultados ofrece, y es la más recomendada por distintos manuales. Este tipo de rotación minimiza el número de variables con saturaciones altas y facilita la interpretación de las componentes.

El resultado de la aplicación de esta rotación es el esperado y da unos componentes principales que pueden ser descritos perfectamente. La descripción de los departamentos se encuentra en el anexo 3.

Comparando las posibilidades de componentes principales que se han obtenido, se decide elegir esta última como la óptima, tanto por cumplir con las teorías estadísticas necesarias, como por el sentido de negocio que se encuentra en ella. Así pues, se prosigue el estudio.

Una vez aplicadas las componentes principales, se procede a la creación de quince nuevas variables: los departamentos a los que representan las grandes variables y provienen de la suma del gasto de los departamentos fuertes que forman a cada componente. Las *querys* que crean a las nuevas variables se encuentran localizadas en el anexo 4.

En el siguiente cuadro se pueden observar los departamentos que han unido las componentes principales y los nuevos nombres de éstas, obtenidos como descriptivos del tipo de productos que se han agrupado en ellas.

CESTAS DE LA COMPRA		
COCINAS	MEDIA SOLUTION AND LIVING ROOM TABLES	DINING Y SOFAS
Knobs and handles	Mirrors	Store and organise furniture
Kitchen appliances	Living room tables	Dining tables
Kitchen taps, sinks and sink accessories	Store and organise furniture	Dining seating
Kitchen worktops	Media solutions & accessories	Freestanding kitchens and kitchen furni
Kitchen fixed interiors	DECORACION OUTDOOR Y TELAS POR METROS	Living room seating
Kitchen fronts	Green decoration	ILUMINACION Y WINDOWS SOLUTION
Kitchen cabinets	Home decoration	Bathroom furniture
EATING AND COOKING	Outdoor	Light sources and accessories
Dining and serving	Cushions, throws and chairpads	Integrated lighting range
Glassware	Wall decoration	Lamps
Cutlery	Fabrics and accessories	Window solutions
Cookware	BAÑOS Y LAVANDERIA	NIÑOS
Kitchen tools	Kitchen accessories	Baby
Storing and washing	Bathroom furniture	Children's storage
Coffee and tea	Laundry, cleaning and sorting	Children 3-7 years
Set the table	Small storage	Play
TEXTILES	Secondary Storage Range	Children's small storage
Bedlinen	Clothes and shoes organisation	WORK
Quilts and pillows	MUEBLES DE DORMITORIO	Store and organise furniture
Bath textiles	Beds	Workspaces
Cushions, throws and chairpads	Wardrobes	drawest desk
FAMILY	Chests & other furniture	seat work
batteries	Mattresses and accessories	ALFOMBRAS O COMPL DE BAÑO
Children's storage furniture	SEASONAL LIGHTING	Bathroom organisers
PAPELERIA Y LIBROS	Kitchen taps, sinks and sink accessories	Home furnishing rugs
child books	Seasonal lighting	

Figura 14. Descripción de componentes principales y departamentos que las forman.

5.2. Tratamiento de outliers

Una vez que se han concluido los pasos anteriores, es el momento de la clasificación de los individuos para reducir su cuantía en grupos más manejables. Para conseguir esto, se aplicarán algoritmos de conglomerados. No obstante, para evitar ruido se procede a la búsqueda y separación del estudio de los *outliers*.

Con este objetivo, se enfrentan las variables creadas (procedentes de los quince componentes) dos a dos para ver gráficamente a los "*raros*" y apartarlos. Esto se hace ya que son clientes con comportamientos extremos que provocarían efectos negativos a la hora de entrenar modelos o de encontrar perfiles similares.

Además, esto servirá para encontrar las "*tarjetas tienda*", es decir, aquellas tarjetas que algunos trabajadores de las tiendas IKEA tienen en caja y utilizan para pasarlas cuando el cliente ha olvidado la suya o no le funciona. Este tipo de tarjetas generan una gran problemática a la hora de realizar estudios, ya que aumentan mucho las

medias de gasto y generan ruido. Así pues, se obtendrán dos trabajos en uno, gracias a la depuración de la base de datos.

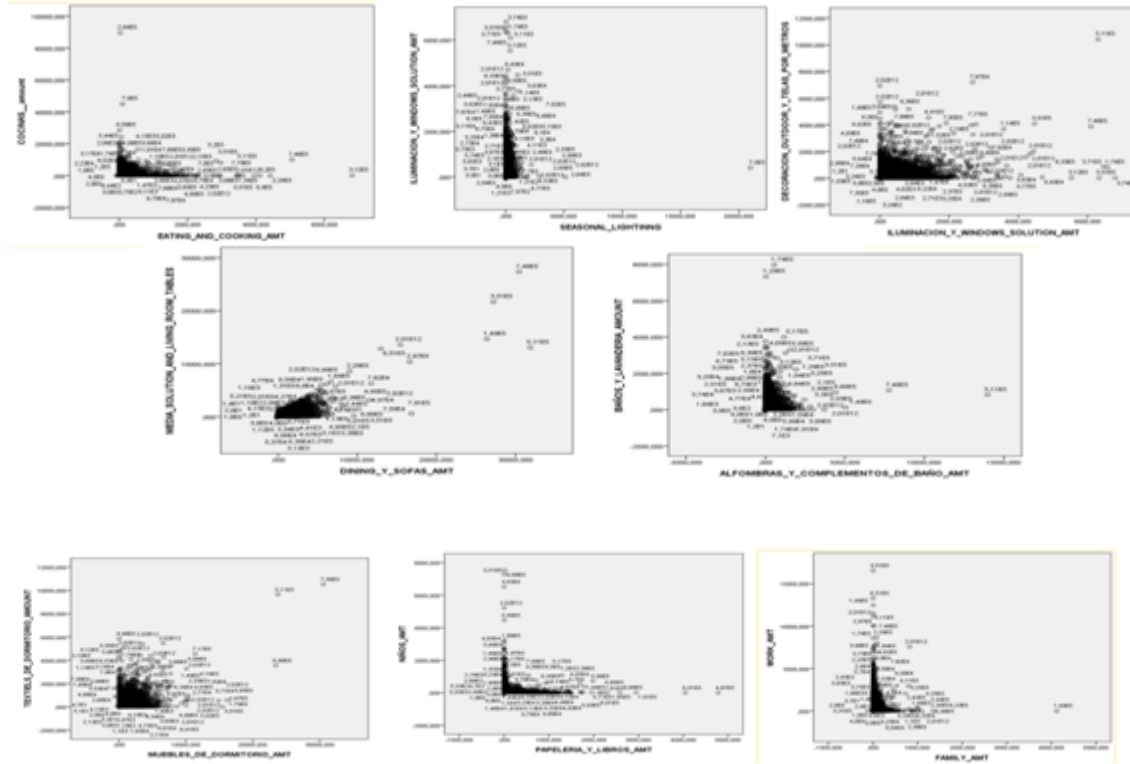


Figura 14. Gráficos de dispersión del gasto de los hogares en los grandes grupos. Detección de outliers.

En las nubes de puntos obtenidas, se observan hogares con valores muy exagerados en comparación con el resto. Por ello se establecen los siguientes filtros en las variables para aislar a esos individuos.

COCINAS__amount < 4000 and *EATING_AND_COOKING_AMT* < 40000 and
TEXTILES_DE_DORMITORIO_AMOUNT < 8000 and *MUEBLES_DE_DORMITORIO_AMT* < 20000 and *MEDIA_SOLUTION_AND_LIVING_ROOM_TABLES* < 10000 and
DINING_Y_SOFA_AMT < 20000 and *BAÑOS_Y_LAVANDERIA_AMOUNT* < 6000 and
ALFOMBRAS_Y_COMPLEMENTOS_DE_BAÑO_AMT < 6000 and
ILUMINACION_Y_WINDOWS_SOLUTION_AMT < 5000 and *SEASONAL_LIGHTNING* < 10000
and *NIÑOS_AMT* < 4000 and *PAPELERIA_Y_LIBROS_AMT* < 3000 and
DECORACION_OUTDOOR_Y_TELAS_POR_METROS < 6000 and *WORK_AMT* < 10000 and
FAMILY_AMT < 2000

Figura 15. Condicionantes para la eliminación de los hogares outliers.

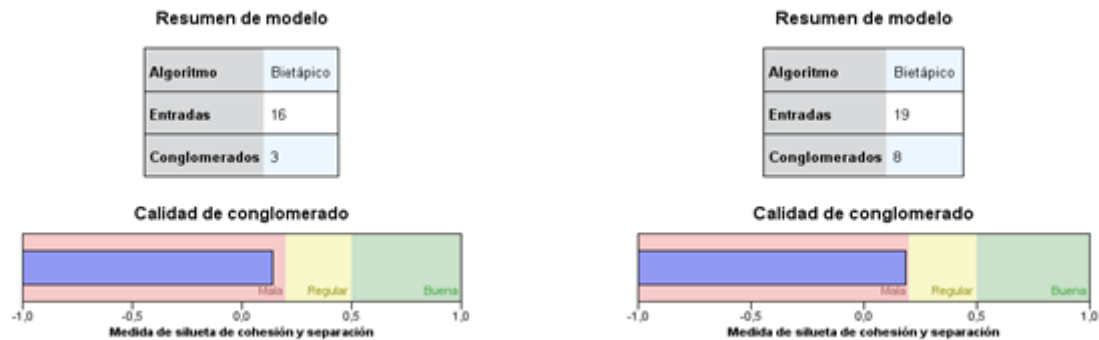
Con esto se consigue depurar la base de datos y, aunque sigue habiendo comportamientos extraños, estos se encuentran dentro de unos cánones. Al ser el objetivo reclasificar a todos los individuos que sea posible, se da por concluido el proceso de filtrado de *outliers*.

5.3. Agrupación de los hogares IKEA Family

Una vez creadas las quince variables procedentes de componentes principales, se busca crear una segmentación para los clientes. Debido a que no se sabe la cantidad de grupos que pueden salir, ni hay un objetivo claro, lo óptimo sería aplicar un algoritmo jerárquico. Sin embargo, esto no es viable con la cantidad de datos que se maneja. Por lo que se aplica el algoritmo k medias que es un no jerárquico, y el bietápico como apoyo.

El algoritmo k medias es el más importante, parte de un vector de medias denominadas centroides que son establecidas (en este caso) por el propio algoritmo, y mediante pruebas sucesivas, contrasta el efecto que sobre la varianza residual tiene la asignación de cada uno de los casos a cada grupo. Como comprueba los distintos casos secuencialmente para ver su influencia individual, el orden puede afectar a su cálculo, sin embargo, sigue siendo el algoritmo que mejor funciona, y por tanto, es el motivo de su elección en este estudio.

Para segmentar a los clientes, hay que tener en cuenta que un cliente debe de ser evaluado en función de su gasto total y por departamento (ambas a la vez), por lo que si se utiliza el gasto por departamento, no es una buena comparación. En cambio, si se hace



porcentualmente, se obtiene la imagen por departamento pero no en total. Es por esto, que se realizaron distintas pruebas en las que se utilizaban las variables porcentuales y de gasto por separado y conjuntamente. Se obtienen resultados válidos, sin embargo, se observó que con la variable gasto hay grandes diferencias por intervalo total de gasto, encontrando clientes que porcentualmente serían identificados en el mismo clúster pero que por cantidad no deberían juntarse, ya que su cuantía es muy distinta. Esto indica, entre otras cosas, quienes son clientes de accesorios y quienes son clientes de muebles.

Además, el algoritmo bietápico señala que la calidad de las clasificaciones obtenidas con ambos tipos de variables no es la mejor.

Figura 16. -izquierda; Algoritmo bietápico utilizando solo variables porcentuales

-derecha; Algoritmo bietápico utilizando solo variables de gasto.

Debido a que, tratando la base de datos completa a la vez, los algoritmos de segmentación no encuentran divisiones correctas de los clientes, se procede a hacer una división de la base de datos manual por calidad de cliente utilizando los cuatro tipos de tipología de cliente que divide a los miembros IKEA Family por gasto y calidad de cliente en; tipología A, B, C y D.

Una vez hecho esto, dentro de cada segmento de individuos similares se procede a encontrar los más parecidos entre sí, y agruparlos en segmentos menores, así pues, se aplican dentro de cada tipología los algoritmos de conglomerados.

Se realiza la aplicación del k medias en cada grupo por lo ya explicado, utilizando las variables porcentuales de gasto en cada departamento por hogar, ya que se busca encontrar patrones de compra de los clientes en función de su tipología.

Puesto que las variables porcentuales están en escalas comparables, no es necesario realizar la tipificación.

5.3.1. Segmentación por tipología de cliente

Se trata de la clasificación de todos los registros de la base de datos en grupos. Estos segmentos deberán tener dentro a los registros más similares entre sí y ser lo más dispares posibles de los registros que forman al resto de grupos. Este tipo de análisis requiere la especificación previa de la cantidad de grupos que se desea obtener.

5.3.1.1. Conglomerados en la tipología A

Se establecen ocho grupos en el cálculo del k medias para comprobar cómo se separan los clientes. El resultado es que ocho clústeres es una cantidad demasiado grande de número de grupos por segmento. Sin embargo, el objetivo es la localización de grupos con posibilidad de unión. Es por esto que se realizan pruebas de cantidades en función de lo que indica el sentido de negocio, la cercanía de los centroides de los diferentes clústeres y lo que indica el algoritmo bietápico.

Centros de los conglomerados finales								
	Conglomerado							
	1	2	3	4	5	6	7	8
X_COCINAS_	69,106	,661	1,620	1,526	1,926	,629	2,099	15,529
X_EATING_AND_COOKING	2,132	1,900	2,411	3,489	3,631	2,500	1,699	3,533
X_TEXTIELS_DE_DORMITORIO	1,980	2,053	3,851	4,843	4,853	3,445	2,451	4,904
X_MEDIA_SOLUTION_AND_LIVING_ROOM_TABLES	1,841	24,499	10,769	9,456	18,423	3,314	2,813	4,776
X_DECORACION_OUTDOOR_Y_TELAS_POR_METROS	1,895	2,583	2,640	4,356	5,091	3,286	2,008	4,253
X_BAÑOS_Y_LAVANDERIA	3,212	3,029	3,319	4,033	5,910	2,409	3,579	16,219
X_MUEBLES_DE_DORMITORIO	4,533	2,580	34,618	9,546	10,151	3,025	72,208	12,643
X_DINING_Y_SOFAS	4,797	30,732	23,743	44,086	20,048	71,783	4,043	9,330
X_ILUMINACION_Y_WINDOWS_SOLUTION	2,867	2,969	3,882	5,208	6,307	2,933	3,153	17,346
X_NIÑOS	1,076	1,244	1,155	1,623	2,343	1,340	1,221	2,020
X_WORK	2,207	25,672	9,789	8,470	17,554	2,863	2,819	5,004
X_ALFOMBRAS_Y_COMPLEMENTOS_DE_BAÑO	,941	1,484	1,683	2,726	2,851	1,942	1,096	2,768
X_FAMILY	,307	,327	,268	,368	,512	,368	,289	,496
X_SEASONAL_LIGHTINNG	3,093	,254	,242	,256	,379	,153	,514	1,159
X_PAPELERIA_Y_LIBROS	,014	,013	,010	,014	,022	,012	,007	,021

Distancias entre los centros de los conglomerados finales								
Conglomerado	1	2	3	4	5	6	7	8
1		80,213	77,262	79,163	73,078	95,894	95,284	58,105
2	80,213		39,006	27,766	17,765	51,554	80,970	44,653
3	77,262	39,006		32,475	27,481	58,421	43,784	36,033
4	79,163	27,766	32,475		27,334	29,870	75,021	41,766
5	73,078	17,765	27,481	27,334		56,614	67,873	29,655
6	95,894	51,554	58,421	29,870	56,614		96,866	68,018
7	95,284	80,970	43,784	75,021	67,873	96,866		64,377
8	58,105	44,653	36,033	41,766	29,655	68,018	64,377	

Figura 17. Centro de los conglomerados y distancia de los centroides. Tipología A

En este caso, en el primer cuadro se observan los departamentos que se unirían en cada conglomerado (marcados en azul los importantes en cada uno de los grupos). En el segundo cuadro, se muestran las distancias de los centroides de los ocho conglomerados, los de menor distancia deberán de unirse para simplificar el estudio (marcados en color).

Se decide mantener tres grupos, los cuales se caracterizan por ser compradores de salón y work, habitación y sofá y cocinas.

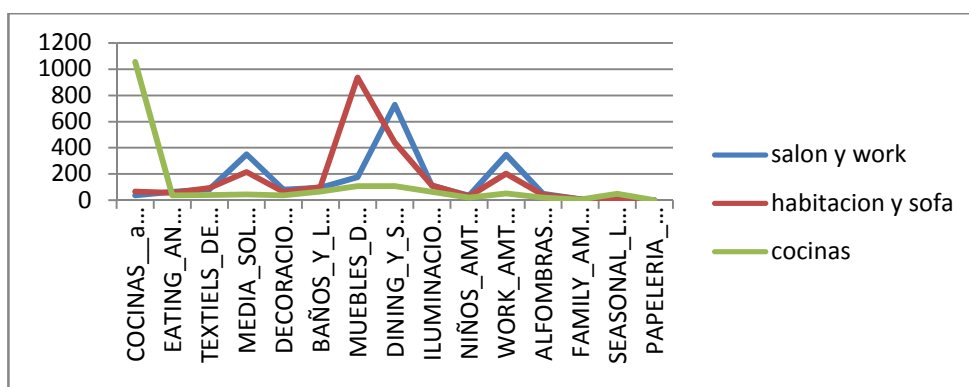


Figura 18. Cantidad de gasto por departamento y clase de cliente de la tipología A

Siendo las siguientes cantidades de hogares las que están en cada grupo:

N de casos en cada conglomerado			%
Conglomerado	1	6696	12%
	2	36937	64%
	3	13664	24%
Válidos		57297	100%
Perdidos		0	0%

Figura 19. Cantidad de hogares por grupo en la tipología A

Esta selección que corrobora el algoritmo bietápico es la más eficiente y, por tanto, es la elegida para este grupo.

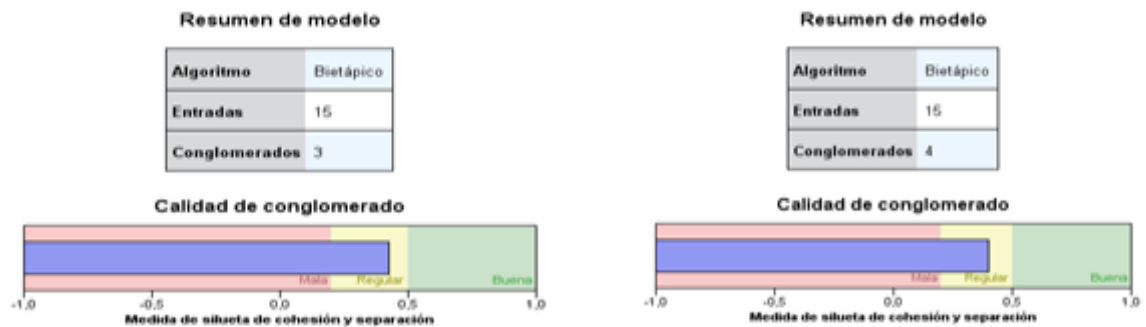


Figura 20. Algoritmo bietápico 3 y 4 nodos para tipología A

5.3.1.2. Conglomerados en la tipología B

Se realizan de nuevo las mismas pruebas con el k medias, comenzando con la selección posible de ocho clústeres distintos para proceder posteriormente a unir los que tengan sentido y su centroide sea cercano.

Centros de los conglomerados finales								
	Conglomerado							
	1	2	3	4	5	6	7	8
X_COCINAS_	28,747	,656	,379	,934	,763	,789	1,081	,872
X_EATING_AND_COOKING	5,703	3,847	2,112	4,423	10,024	2,454	5,810	4,843
X_TEXTIELS_DE_DORMITORIO	5,865	4,298	2,094	7,222	16,874	5,284	5,776	5,571
X_MEDIA_SOLUTION_AND_LIVING_ROOM_TABLES	4,224	4,403	28,113	5,909	4,690	2,540	4,520	10,057
X_DECORACION_OUTDOOR_Y_TELAS_PO_R_METROS	5,997	4,338	2,737	5,414	14,124	2,887	7,337	6,857
X_BAÑOS_Y_LAVANDERIA	7,048	3,616	3,101	5,679	8,848	3,201	20,127	7,974
X_MUEBLES_DE_DORMITORIO	10,598	4,996	2,208	34,645	5,930	67,580	4,718	6,731
X_DINING_Y_SOFAS	10,080	59,947	27,242	14,988	10,385	4,292	7,651	17,014
X_ILUMINACION_Y_WINDOWS_SOLUTION	7,047	4,437	2,722	6,396	9,097	3,486	28,083	7,418
X_NIÑOS	3,738	1,984	1,387	3,497	6,204	2,069	3,732	4,004
X_WORK	5,249	4,269	25,979	6,880	4,677	3,291	4,790	23,409
X_ALFOMBRAS_Y_COMPLEMENTOS_DE_B_AÑO	3,099	2,534	1,316	3,085	6,637	1,476	4,753	3,968
X_FAMILY	,840	,477	,400	,630	1,255	,406	,964	,860
X_SEASONAL_LIGHTINNG	1,733	,184	,194	,277	,445	,231	,624	,387
X_PAPELERIA_Y_LIBROS	,032	,015	,016	,021	,046	,014	,034	,036
Distancias entre los centros de los conglomerados finales								
Conglomerado	1	2	3	4	5	6	7	8
1		57,804	47,469	37,279	32,239	64,223	37,756	34,772
2	57,804		46,113	54,132	53,236	83,821	59,921	47,830
3	47,469	46,113		46,119	42,863	77,334	48,943	23,545
4	37,279	54,132	46,119		33,018	35,450	40,566	32,956
5	32,239	53,236	42,863	33,018		65,346	26,335	25,444
6	64,223	83,821	77,334	35,450	65,346		70,042	66,299
7	37,756	59,921	48,943	40,566	26,335	70,042		32,338
8	34,772	47,830	23,545	32,956	25,444	66,299	32,338	

Figura 21. Centro de los conglomerados y distancia de los centroides. Tipología B

Debido a la cantidad de individuos que hay en cada clúster como por la unión de centroides, queda claro que hay que disminuir el número de conglomerados. La cantidad más eficiente de todas, debido a la predicción del algoritmo bietápico, del sentido de negocio y de la unión de centroides, es la de cuatro grupos.

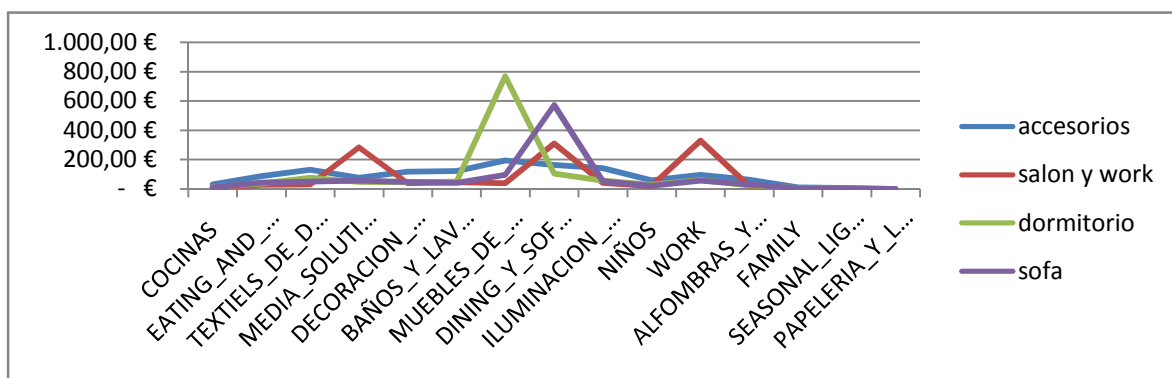


Figura 22. Cantidad de gasto por departamento y clase de cliente de la tipología B

Estos tienen un perfil de compradores de accesorios, de salón y *work*, de dormitorio y de sofá.

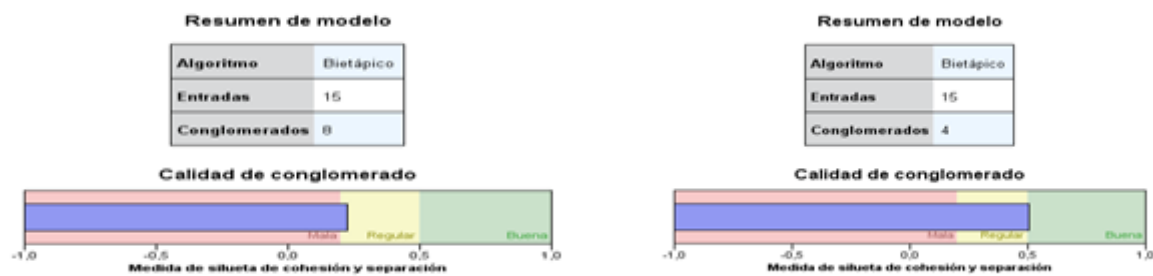


Figura 23. Algoritmo bietápico 3 y 4 nodos para tipología B

Las cantidades de hogares en cada clúster son las siguientes

N de casos en cada conglomerado			%
Conglomerado	1	27529	32%
	2	9811	12%
	3	31432	37%
	4	16143	19%
Válidos		84915	100%
Perdidos		11	0%

Figura 24. Cantidad de hogares por grupo en la tipología B

5.3.1.3. Conglomerados en la tipología C

En el grupo de hogares pertenecientes a la tipología C se utiliza el algoritmo k medias y bietápico nuevamente. Se aplican ocho segmentos distintos y se continúa disminuyendo este número progresivamente a través de clústeres similares.

Centros de los conglomerados finales								
	Conglomerado							
	1	2	3	4	5	6	7	8
X_COCINAS_	3,487	1,102	1,355	,354	,390	,758	1,223	,596
X_EATING_AND_COOKING	9,471	4,765	27,929	7,018	5,063	5,148	8,014	3,788
X_TEXTIELS_DE_DORMITORIO	24,838	4,229	5,544	4,765	4,159	4,852	6,221	5,013
X_MEDIA_SOLUTION_AND_LIVING_ROOM_TABLES	2,596	22,988	2,149	1,440	2,804	2,897	3,337	2,516
X_DECORACION_OUTDOOR_Y_TELAS_PO R_METROS	21,509	5,814	9,760	6,568	5,470	5,504	8,936	4,008
X_BAÑOS_Y_LAVANDERIA	7,371	6,489	26,091	5,617	5,770	4,317	12,567	4,429
X_MUEBLES_DE_DORMITORIO	2,603	4,047	1,661	2,549	1,610	2,626	2,563	63,875
X_DINING_Y_SOFAS	2,848	18,958	2,108	1,720	3,134	60,554	3,026	2,949
X_ILUMINACION_Y_WINDOWS_SOLUTION	5,777	4,907	4,863	3,646	4,973	4,295	37,958	3,911
X_NIÑOS	5,431	3,170	5,920	59,119	2,568	2,472	4,684	3,158
X_WORK	2,754	18,954	2,326	1,649	59,992	2,956	3,100	2,844
X_ALFOMBRAS_Y_COMPLEMENTOS_DE_B AÑO	7,436	2,951	5,536	2,299	2,031	2,292	5,290	1,794
X_FAMILY	3,199	1,291	4,112	3,029	1,825	1,077	2,470	,939
X_SEASONAL_LIGHTINNG	,588	,294	,544	,156	,169	,216	,535	,153
X_PAPELERIA_Y_LIBROS	,093	,041	,102	,072	,041	,036	,077	,026

Distancias entre los centros de los conglomerados finales								
merado	1	2	3	4	5	6	7	8
1		40,856	34,833	59,686	63,513	63,743	39,761	67,444
2	40,856		44,281	64,852	48,483	48,975	45,346	67,276
3	34,833	44,281		60,940	65,673	66,797	41,035	70,643
4	59,686	64,852	60,940		81,335	81,760	64,923	83,185
5	63,513	48,483	65,673	81,335		80,963	66,434	84,567
6	63,743	48,975	66,797	81,760	80,963		67,446	84,113
7	39,761	45,346	41,035	64,923	66,434	67,446		71,037
8	67,444	67,276	70,643	83,185	84,567	84,113	71,037	

Figura 25. Centro de los conglomerados y distancia de los centroides. Tipología C

Realizando pruebas de cantidades, finalmente se eligen dos clústeres, mostrando una gran diferenciación entre los hogares compradores de productos del departamento de niños.

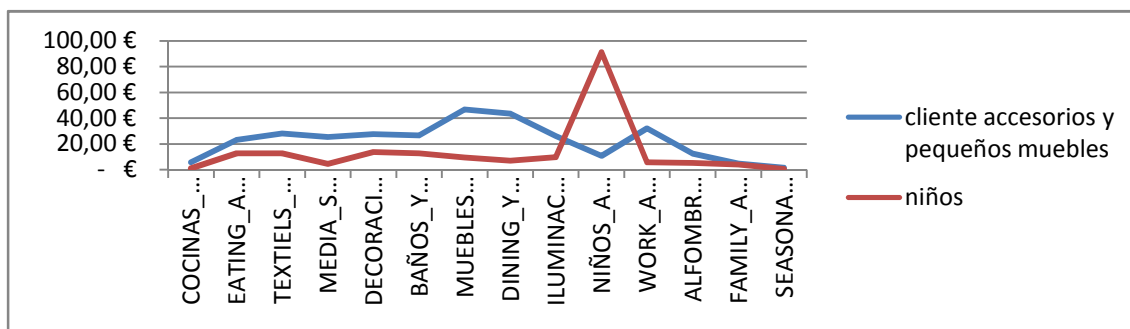


Figura 26. Cantidad de gasto por departamento y clase de cliente de la tipología C

El algoritmo bietápico apoya la elección de dos segmentos frente al resto.

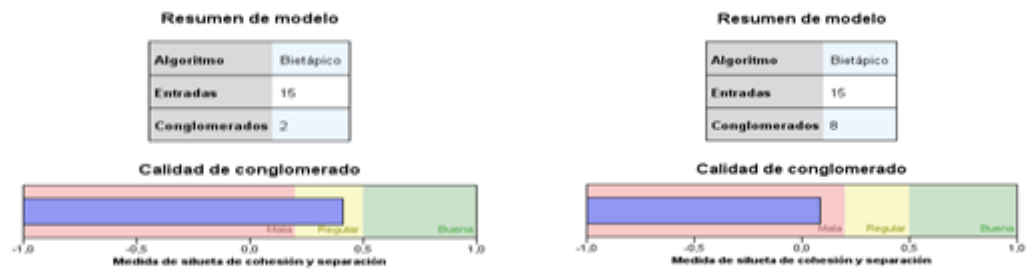


Figura 27. Algoritmo bietápico 3 y 4 nodos para tipología C

Las proporciones de hogar en cada segmento son representativas, por lo que se mantiene esta selección.

N de casos en cada conglomerado			%
Conglomerado	1	444431,000	89%
	2	54116,000	11%
Válidos		498547,000	100%
Perdidos		57,000	0%

Figura 28. Cantidad de hogares por grupo en la tipología C

5.3.1.4. Conglomerados en la tipología D

Los hogares de la tipología D serán divididos en distintos segmentos. Se comienza con una gran cantidad para ir reduciendo progresivamente y seleccionar la cantidad óptima. El k medias indica, que la cantidad recomendada es de tres clústeres.

Centros de los conglomerados finales			
	Conglomerado		
	1	2	3
X_COCINAS_	0,000	0,000	0,000
X_EATING_AND_COOKING	8,324	7,588	72,254
X_TEXTIELS_DE_DORMITORIO	21,700	6,563	4,332
X_MEDIA_SOLUTION_AND_LIVING_ROOM_TABLES	0,000	0,000	0,000
X_DECORACION_OUTDOOR_Y_TELAS_POR_METROS	25,289	7,843	9,017
X_BAÑOS_Y_LAVANDERIA	13,773	4,142	5,246
X_MUEBLES_DE_DORMITORIO	0,000	0,000	0,000
X_DINING_Y_SOFAS	0,000	0,000	0,000
X_ILUMINACION_Y_WINDOWS_SOLUTION	16,618	5,788	3,518
X_NIÑOS	1,984	,380	,603
X_WORK	0,000	0,000	0,000
X_ALFOMBRAS_Y_COMPLEMENTOS_DE_BAÑO	1,016	65,242	,740
X_FAMILY	10,843	2,342	4,124
X_SEASONAL_LIGHTINNG	,278	,041	,067
X_PAPELERIA_Y_LIBROS	,176	,070	,098

Figura 29. Centro de los conglomerados y distancia de los centroides. Tipología D

El bietápico confirma que es una buena cantidad de grupos.



Figura 30. Algoritmo bietápico 3 nodos para tipología D

Esta división divide a los hogares en clientes de decoración, de accesorios y de alfombras.

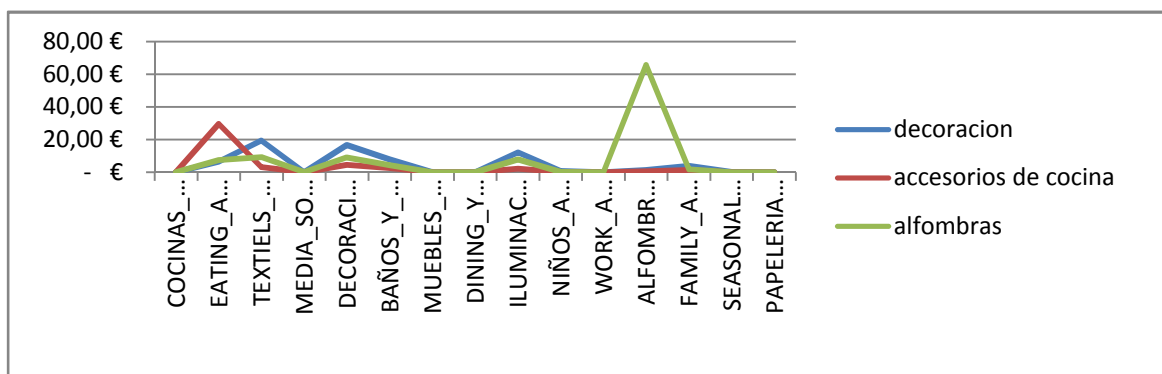


Figura 31. Cantidad de gasto por departamento y clase de cliente de la tipología D

Las cantidades de cada *clúster* son eficientes, y por tanto, se mantiene esta selección.

N de casos en cada conglomerado			
Conglomerado	1	68437,000	75%
	2	4133,000	5%
	3	18754,000	21%
Válidos		91324,000	100%
Perdidos		973,000	1%

Figura 32. Cantidad de hogares por grupo en la tipología D

5.3.2. Segmentación por conducta de cliente

Obtenidos ya todos los clústeres significativos en cada tipología, se ha conseguido que las formaciones de grupos tengan en cuenta la calidad del cliente, su gasto y sus gustos. Ahora es el momento de conocer si estos grupos son simplificables y analizar la posible la unión de los más parecidos según su comportamiento de consumo.

Debido a que la base de datos tiene la posibilidad de reducirse a través de agregados a una matriz formada por quince componentes principales y doce clústeres, es viable aplicar el análisis jerárquico para comprobar si une distintos grupos y la cantidad con la que se queda. Esto se hace ya que el siguiente paso es aplicar árboles de decisión dentro de cada clúster y si se consigue reducir la cantidad, disminuye la complejidad del proceso de reclasificación de los individuos.

El análisis jerárquico, se enfrenta a doce clústeres con características de compra propias muy distintas entre algunos grupos y similares con otros. Se busca encontrar la unión de estos grupos similares. Para ello, se han tenido en cuenta los centroides de cada uno de los distintos segmentos.

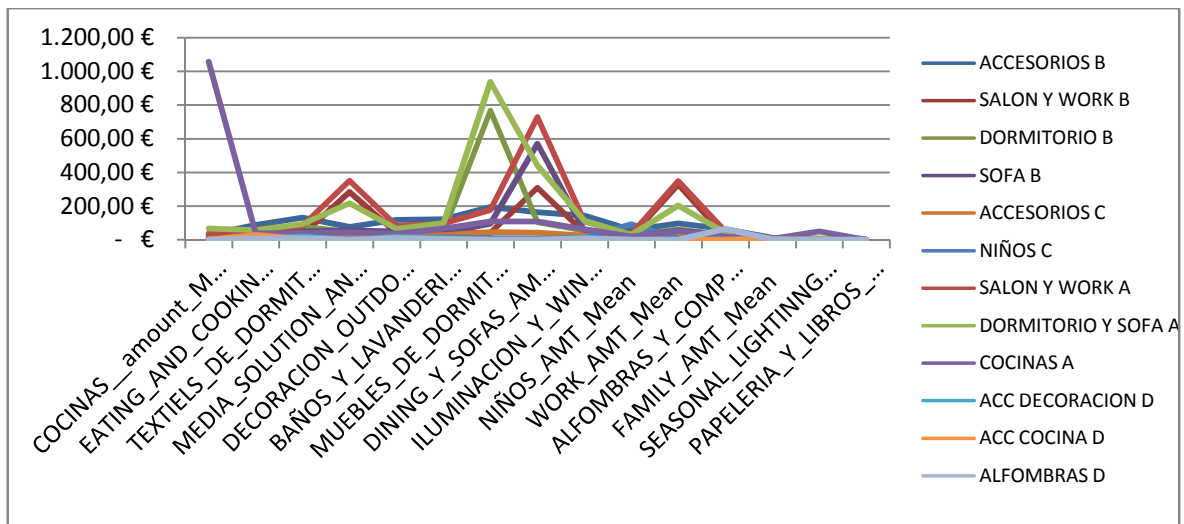


Figura 33. Cantidad de gasto por departamento y clase de cliente de todas las tipologías

Este algoritmo ha agrupado los doce grupos en cuatro categorías distintas. Por lo que la matriz pasaría a ser de quince componentes por cuatro clústeres.

Estos cuatro clústeres son los siguientes;

- **Clientes de accesorios**
 - Accesorios de decoración tipología D
 - Accesorios de cocina tipología D
 - Alfombras tipología D
 - Accesorios de cocina tipología D
 - Niños C
 - Accesorios B
- **Clientes de salón**
 - Salón y *work* B
 - Sofá B
 - Salón y *work* A
- **Clientes de dormitorio**
 - Dormitorio B
 - Dormitorio y sofá A
- **Clientes de Cocina**
 - Cocina A

Se considera una buena agrupación, ya que es la misma que indicaba el sentido de negocio.

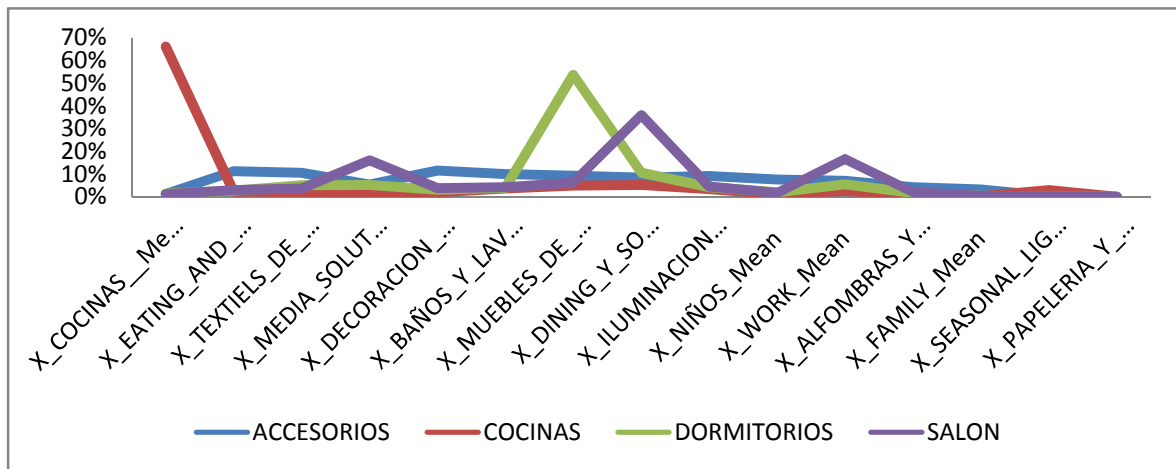


Figura 34. Porcentaje de gasto por departamento de los cuatro *clústeres* obtenidos

Se observan claramente en la gráfica los perfiles de cliente según su consumo por departamento.

Las cantidades en cada grupo son muy diferentes, sin embargo, es algo normal, ya que la gran mayoría de consumidores son *clientes de accesorios* en el primer año, siendo los más extraños los *clientes estrella*, aquellos que tienen una confianza plena en la marca, y por ello, compran la cocina completa IKEA.

CLUSTER4	Cantidad de HHID	% Total
ACCESORIOS	621676	85%
SALON	62692	9%
DORMITORIOS	41019	6%
COCINAS	6696	1%
NULOS	1041	0%
TOTAL	733124	100%

Figura 35. Cantidad de hogares por tipo de cliente

En el dendograma se muestra la unión de los distintos grupos a través del algoritmo jerárquico. En él, se observa claramente la unión entre grupos de clientes de menor categoría, es decir, de menor gasto y confianza en la marca IKEA como lo son los clientes de accesorios, seguidos de los de muebles de salón, sofás y *work*, y como cliente estrella, los clientes de cocinas. Se ha decidido elegir las uniones del segundo nivel del dendograma, ya que se ha considerado una cuantía con sentido de negocio y manejable para realizar los árboles de clasificación.

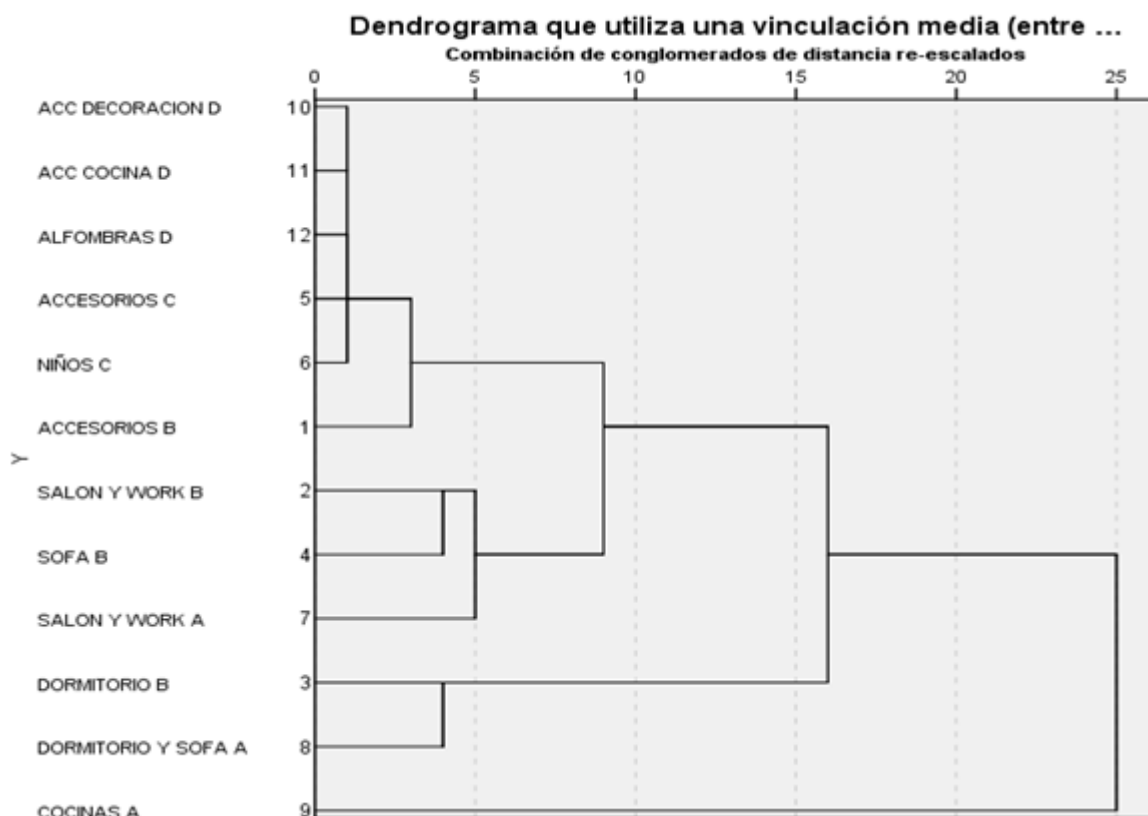


Figura 36. Dendrograma procedente del análisis jerárquico

Para asegurar que la elección tomada es la correcta, se aplica también el algoritmo bietápico, que indica que la calidad de la segmentación supera los niveles y es correcta la calidad de los conglomerados.



Figura 37. Resultado análisis bietápico 4 nodos

Una vez clasificados los individuos, se vuelve a trabajar con *SPSS Modeler* creando una nueva variable del fichero obtenido con *SPSS Statistics*, llamada *CLÚSTERS4*,

donde los clientes pertenecientes a los distintos *clústeres* son identificados como "ACCESORIOS", "SALÓN", "DORMITORIOS", "COCINAS" y "NULOS".

5.4. Resultados del análisis. Clasificación de los individuos

Gracias a esta diferenciación del tipo de clientes recién obtenida, se realiza un filtrado para trabajar con cada uno de los *clústeres* individualmente. Es decir, se obtienen cuatro tablas distintas, una por tipo de cliente, a excepción de los nulos que apartamos del estudio ya que son una pequeña cantidad y muy distintos unos de otros como para tratarlos como iguales.

Por separado, en cada uno de los cuatro ficheros se calculan las nuevas variables con el objetivo de clasificar al hogar según su gasto. Teniendo en cuenta que ahora los clientes son similares tanto en gasto como en porcentaje de gasto total, es coherente crear la variable clientes por departamento que indiquen si un individuo *no es cliente*, es *cliente de baja calidad* o es un *cliente estrella* de cada una de las variables creadas a partir de las componentes principales. Esto se hace a través del cálculo de la media de gasto de cada hogar en cada variable. Con esto, se establecen tres categorías distintas;

- '0'; no es cliente del departamento.
- '1'; es cliente no característico de ese departamento (gasto inferior a la media).
- '2'; es un cliente estrella del departamento (gasto igual o superior a la media).

Con el objetivo de que el árbol clasifique mejor, se transforman las variables de cliente de tipo continuo a tipo nominal. Esto se hace a través del nodo *rellenar* con la opción de *cambiar en la misma variable* con la petición de *"to string (@FIELD)"*.

Así pues, se ha obtenido un archivo por *clúster*, donde hay catorce variables más y la variable de edad en intervalos, ya que se realizaron pruebas y los árboles son más sencillos y sin problemas de ajuste si se realizan intervalos de edad dentro de los diferentes *clústeres*.

Así, ahora, los archivos con los que se trabajará son los siguientes:

- SALÓN CON EDAD POR INTERVALOS PRUEBA
- COCINA CON EDAD POR INTERVALOS PRUEBA
- DORMITORIO CON EDAD POR INTERVALOS PRUEBA
- ACCESORIOS CON EDAD POR INTERVALOS PRUEBA

Los cuales tienen las variables de clasificación de cliente con valores '0', '1' y '2.'

PAPELERIA	→	PAPELERIA
SEASONAL	→	SEASONAL
FAMILY	→	FAMILY
ALFOMBRAS	→	ALFOMBRAS
WORK	→	WORK
NIÑOS	→	NIÑOS
ILUMINACION	→	ILUMINACION
SOFAS	→	SOFAS
DORMITORIO	→	DORMITORIO
BAÑOS	→	BAÑOS
DECORACION	→	DECORACION
LIVING_ROOM	→	LIVING_ROOM
COOKING	→	COOKING
COCINAS	→	COCINAS

Figura 38. Variables reclasificadas como nominales.

Preparados los ficheros en *SPSS Statistics*, se procede a la aplicación de árboles *CHAID exhaustivos* en los cuatro *clústeres*. Se impone como variable dependiente la variable *hijos*, marcando como objetivo la respuesta 'sí'. Se utilizan como variables independientes las recién creadas de clientes, se introduce también la variable *EDAD* que es la que tiene dividida la edad de los individuos en intervalos.

Se aplica al árbol un coste de dos puntos si clasifica en *hijos* = 'no' al hogar que tiene marcado en el dato real la opción de *hijos* = 'sí', ya que el árbol busca reducir el error por lo que tenderá a clasificar los de menor cuantía ('sí' en *hijos*) en el grupo de mayor cuantía ('no' en *hijos*).

En Portugal hay 3.226.371 familias, de estas, 2.094.732 (65%) tienen hijos, y que de estos al menos uno sea menor de 15 años hay 1.105.066 (34%) familias. Puesto que IKEA considera que una familia tiene hijos si alguno de ellos tiene 12 o menos años, con conseguir un 25% se considera un buen resultado.

En la base de datos actual se tiene un 18% de los clientes reconocidos como hogares con hijos (135.476 hogares). Se busca conseguir un 25%, es decir, 183.281 hogares marcados como con hijos. Esto se consigue a través de la imposición de coste dos de una mala clasificación por parte del árbol.

Se construyen los cuatro árboles de decisión dentro de cada clúster. La importancia de este paso reside en:

- Conocer los departamentos que utiliza el modelo para perfilar al cliente con hijos para futuros refrescos del proyecto.
- Conocer los departamentos que utiliza para aplicarlo al total de vida en los clientes que dicen tener hijos pero que en el primer año de compra no lo han demostrado, motivo por el que el modelo los clasifica en sin hijos erróneamente, para conocer si es un cliente con hijos pero que no compra o un cliente que ha tenido hijos después del primer año de vida de cliente IKEA.
- Obtener el cuadro de errores para la posterior definición de acciones en los hogares pertenecientes a cada tipo de error.

Este último otorgará la clasificación esperada y por la que se ha realizado el estudio, ya que indicará los aciertos, es decir, coincidentes, hogares que indicaron que sí o no tenían hijos en el domicilio y que el modelo predice como tales, y los "*errores*". Los errores de clasificación en los que el modelo prediga *hijos* = '*sí*' y el individuo haya indicado *hijos* = '*no*', serán aquellos padres que denominamos con anterioridad como *padres potenciales*.

Los errores en los que los hogares dicen tener hijos y el modelo indica que no parecen padres, indica que son padres que no están realizando sus compras en IKEA, lo que provoca que haya que separarlos y analizar su vida completa de clientes para conocer si en algún momento han realizado compras que se asimilen al perfil de hogares con hijos. Si es así, significará que el hogar actualizó el campo de hijos después del primer año de cliente a *hijos* = '*sí*', pero en el primer año no tenía niños en el hogar (este posible error se debe a que no hay ninguna variable que indique cuando actualizó el cliente sus datos). Si no es coincidente con el perfil de hogar con hijos en sus años de vida de cliente, significará que tienen un gran potencial de gasto en los departamentos característicos de niños y que

hay que aplicar procesos de reactivación y unión con la marca, ya que son un nicho clave de mercado y con un buen potencial de compra. Los árboles clasificadores se encuentran en el anexo 5.

Todos ellos se caracterizan por la gran importancia en la predicción que marca la edad del individuo, seguida del ser o no cliente del departamento de *family* o niños según la edad media de los hogares. Así, se encuentra que la cantidad de hogares con hijos según la edad hace una campana en la que entre los 38 y 44 años se da el máximo de porcentaje de padres con hijos, siguiendo anteriormente una tendencia de crecimiento y posteriormente de decrecimiento.

Los hogares con niños y con edades medias inferiores a los 38 años están marcados por ser clientes del departamento de niños en primer nivel y los mayores de esta edad, están marcados por el departamento de *family*.

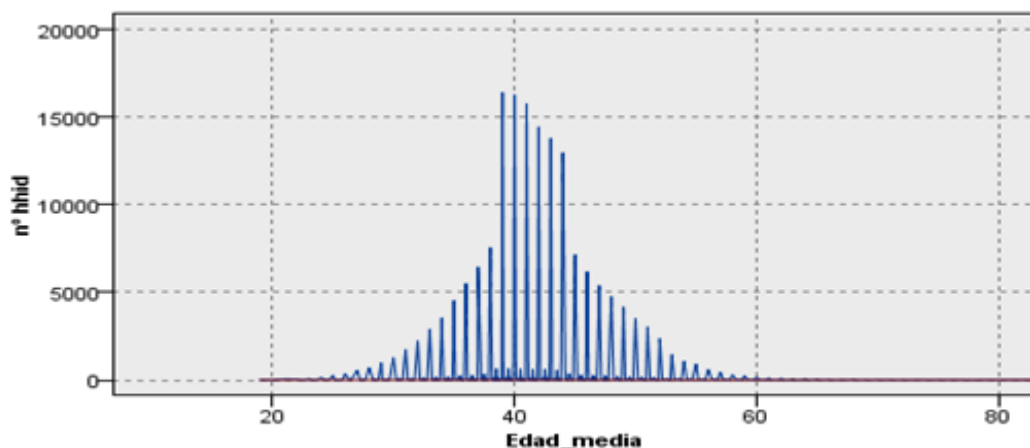


Figura 39. Cantidad de hogares con hijos con o menores de 12 años en la predicción

Si se presta atención a los siguientes niveles de las ramas de los diferentes árboles, en los segmentos de menores de 35 años, los hogares con niños se caracterizan por ser cliente de baños, *family* e iluminación. Entre los 35 a 59 años de edad, los miembros Family se caracterizan por ser cliente de dormitorios, papelería, *family* y *work*, y posteriormente salen comportamientos de compra de niños nuevamente. Esto refleja la correlación positiva de la edad de los niños con la edad de los padres.

MOBILIARIO INFANTIL

Para las personas más importantes del mundo

Los niños están muy ocupados en descubrir quiénes son y el mundo que les rodea, y empezando su camino para ser independientes. Tener los artículos adecuados en casa puede facilitarles ese proceso. Desde que llegan a casa hasta que ya son mayorcitos, nuestros muebles infantiles te ayudan a convertir tu hogar en el mejor patio de recreo.



Bebés (341)

Cunas, Colchones para cunas, Textiles para bebés, Juguetes para bebés, Muebles de almacenaje, Vajillas para niños, Sistema de almacenaje, Almacenaje pequeño, Cambiadores, Iluminación. Niños, Tronas, Seguridad



Niños de 3 a 7 años (469)

Camas de niños, Colchones para niños, Textiles para niños, Mobiliario para niños, Muebles de almacenaje, Sistema de almacenaje, Almacenaje pequeño, Juguetes y Juegos, Sillas junior, Iluminación. Niños, Juego de mesa para niños, Espejos para niños, Seguridad



Niños de 8 a 12 años (276)

Camas de niños 8-12 años, Colchones de niños 8-12 años, Tejidos para niños 8-12 años, Accesorios para niños 8-12 años, Muebles de almacenaje, Sistema de almacenaje, Almacenaje pequeño, Iluminación niños 8-12 años, Escritorios y sillas de niños 8-12 años

Figura 40. Clasificación de la diferenciación de productos de niños por edad

Antes de los 35 años, los hijos son niños que aún no han empezado el colegio, es por eso que las compras relacionadas son aquellas de habituación de los niños al hogar (departamento de niños; cunas, juguetes, adaptadores de asiento...), ser cliente de *family* en esta etapa es lógico, ya que este departamento tiene la gama de productos *PATRULL*, es decir, los complementos de seguridad para los niños. También es coherente que aparezcan como descriptivos el ser clientes de decoración o de iluminación, ya que estos departamentos tienen productos adaptados a estas edades.

Cuando los hogares superan la edad media de 35 años, los niños también tienen un cambio en su vida, y pasan a empezar el colegio, a formarse y con ello a necesitar adaptar su espacio a las nuevas necesidades. En este cambio, el espacio que se dedicaba al descanso y juego del niño ahora pasa dedicarse al almacenaje (de mayor cuantía) para sus cosas (ropa, juguetes, libros...). Se evoluciona de la cuna a la cama y no basta con jugar, el niño tiene que estudiar, por lo que necesita un lugar de estudio (*work*) y necesita su material escolar para empezar sus clases (mochilas (*family*), papelería...). Es por estos cambios que las familias con estas edades medias se caracterizan por ser clientes *de family*, papelería, *work* y habitación.

A partir de los 60 años, se obtienen comportamientos extraños, pasando a ser de nuevo una señal de hogares con niños de corta edad al ser clientes del departamento de niños. Estos hogares no son reales hogares con niños, y pese que hay hogares con hijos tardíos, este comportamiento es extraño y se deduce que este grupo de clientes son abuelos que están comprando muebles y regalos a sus nietos. Sin embargo, puesto que lo que

interesa a IKEA no es conocer si son clientes de cosas para hijos o nietos, los clasificamos como padres ya que están demostrando un comportamiento de hogar con niños.

Con el fin de la clasificación de los individuos en hogares con o sin niños, se procede a cuantificar cada tipo de hogar por clúster.

- Cliente 'ACCESORIOS'

Clasificación			
Observado	Pronosticado		Porcentaje correcto
	no	si	
no	449367	39322	92,0%
si	90852	22949	20,2%
Porcentaje global	89,7%	10,3%	78,4%
Métodos de crecimiento: CHAID exhaustivo			

Figura 41. Tabla de errores del árbol de decisión aplicado en ACCESORIOS

- Cliente 'DORMITORIO'

Clasificación			
Observado	Pronosticado		Porcentaje correcto
	no	si	
no	29739	2716	91,6%
si	5780	1651	22,2%
Porcentaje global	89,1%	10,9%	78,7%
Métodos de crecimiento: CHAID exhaustivo			

Figura 42. Tabla de errores del árbol de decisión aplicado en DORMITORIOS

- Cliente 'SALÓN'

Clasificación			
Observado	Pronosticado		Porcentaje correcto
	no	si	
no	45056	4504	90,9%
si	8595	2799	24,6%
Porcentaje global	88,0%	12,0%	78,5%
Métodos de crecimiento: CHAID exhaustivo			

Figura 43. Tabla de errores del árbol de decisión aplicado en SALÓN

- Cliente 'COCINAS'

Clasificación			
Observado	Pronosticado		Porcentaje correcto
	no	si	
no	4905	412	92,3%
si	911	251	21,6%
Porcentaje global	89,8%	10,2%	79,6%
Métodos de crecimiento: CHAID exhaustivo			

Figura 44. Tabla de errores del árbol de decisión aplicado en COCINAS

6. Conclusiones y estrategias a seguir

Los porcentajes de adaptación de las clasificaciones a los datos reales son buenos (todos mayores del 78%). Sin la corrección de "*costes de error* = 2", la adaptación es mayor, sin embargo, era un error por nuestra parte aceptarlo, ya que el error disminuía a cambio de que el árbol clasificará a la minoría (*hijos* = 'sí') en la mayoría (*hijos* = 'no').

Las cantidades totales de la clasificación son las siguientes:

PRONOSTICO			REALIDAD	
			no	si
	no	Recuento	529.067	106.138
		% del total	75%	15%
	si	Recuento	46.954	27.650
		% del total	7%	4%

Figura 45. Matriz de doble entrada del total de hogares clasificados según las predicciones y su respuesta

- **No padres;** Un 75% de los individuos no tienen hijos y son considerados como tal.
- **Padres;** Un 4% dicen tener hijos y son considerados como padres.
- **Padres sin explotar;** Un 15% dice tener hijos y no es considerado como hogar con hijos por el modelo.
- **Padres potenciales;** Un 7% dicen no tener hijos y son considerados padres por el modelo.

El número total de hogares con niños que se estima que hay en IKEA Family es de (106.138+27.650+46.954) 180.742 de los 709.809 hogares que se han clasificado finalmente. Es decir, un 25% de todos los hogares seleccionados tienen niños con edades inferiores o iguales a los 12 años, proporción buscada desde el inicio del ejercicio. Con lo que, puesto que los datos dan resultados coherentes tanto a nivel matemático como de negocio y demográfico, se establece como buen resultado y se concluye que este es el primer paso para la limpieza de la base de datos de IKEA Family Portugal y el inicio de la puesta en marcha de la personalización de los impactos por tipo de familia.

Este análisis se ha realizado para conseguir establecer resultados coherentes con los datos demográficos del país. Ahora, IKEA Family debe confirmar que la clasificación que se ha creado es la correcta. Por ello, se procederá en el siguiente mes a enviar a los hogares denominados *padres potenciales* un email o un sms, según la forma de contacto que tengan estos hogares disponible, con el que entrarán en un sorteo con un premio. Esto se realiza con el fin de incentivar a los individuos a que actualicen sus datos de perfil, siendo una de las preguntas la referente a ser o no padre, y estando en la primera posición para evitar que contesten a otras preguntas y no a esa. Se introducen más preguntas, en vez de solo la referente a los hijos para evitar posibles problemas como el del caso Target*.

Con los hogares clasificados como *padres sin explotar*, se obtiene el histórico completo de compras de estos hogares (no solo del primer año) a través de un cruce por hogares coincidentes con el archivo *PT_Ventas*. Se reestructuran las variables de departamento como se hizo con la preparación de datos y se establecen nuevamente los departamentos obtenidos de las componentes principales a través de nodos *derive*. Finalmente, se introduce la tipología a los hogares a través de un *nodo fusionar* por *hhidd* coincidentes con el archivo *PT_tipologia*, y se establece un filtro de perfil de cliente en función del tipo de perfil de clientes con niños según la tipología a la que pertenezca, es decir, se introducen *queries* en los que los hogares deben de superar alguna media de gasto en los departamentos importantes para la clasificación de hogares con niños (departamento de niños, *work*, papelería, *family* y dormitorio). En función de la posición de gasto de estos hogares en cada departamento respecto a la media de gasto en estos cuatro grupos, se mantienen en *padres sin explotar* si no superan el filtro o pasan al grupo de *padres* si superan alguna de las condiciones.

*Caso Target; La empresa Target junto con Andrew Pole trato de identificar a las mujeres en sus primeros meses de embarazo a través de sus hábitos de compra. Esta predicción se obtuvo a través de la observación de cambios en el comportamiento de consumo de determinados productos y cantidades de estos (lociones, suplementos dietéticos...). Target con el objetivo de venta cruzar al cliente, envió cupones de descuento de cosas específicas de bebé a una adolescente. El padre de esta, indignado, acusó a la empresa de animar a su hija a quedarse embarazada. Finalmente, descubrió que su hija estaba embarazada.

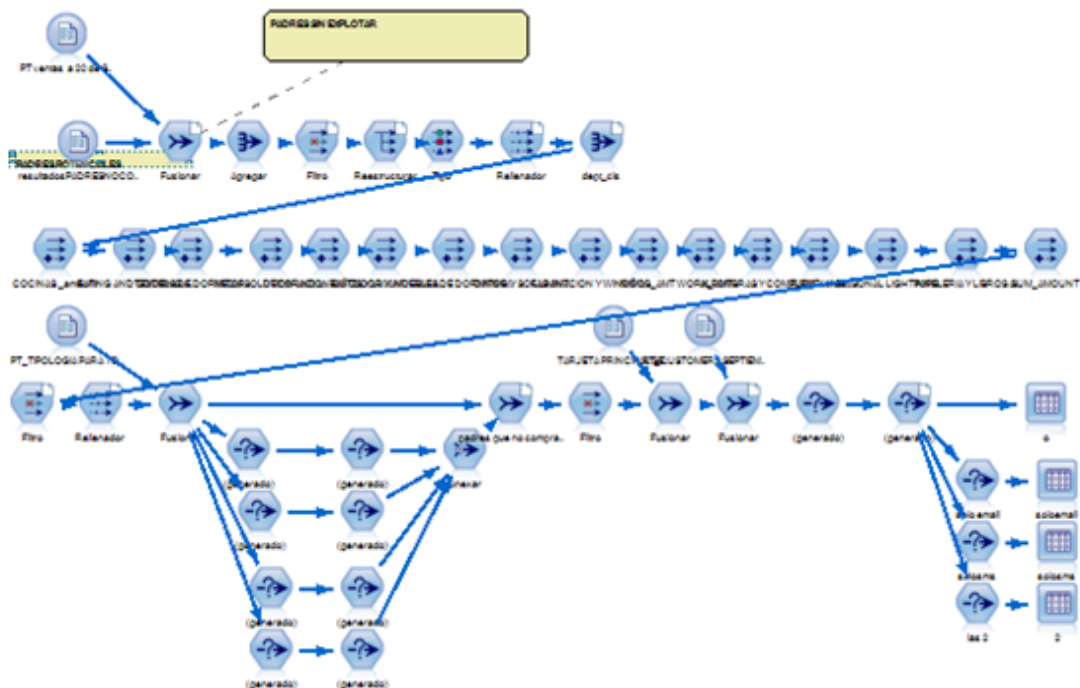


Figura 46. Stream de seguimiento de los hogares a lo largo de su vida que el modelo predice como no padres y ellos dicen serlo

El resultado obtenido es que de los 106.138 hogares, 33.885 han sido reclasificados por el modelo a *padres* a lo largo de su ciclo de vida, lo que indica que han tenido hijos más adelante y han actualizado en su perfil la variable *with children*. 72.253 hogares no han cambiado su clasificación en el modelo, por lo que en su vida de clientes IKEA no han tenido el perfil de cliente con niños. Es por esto que este grupo es el que recibirá acciones específicas para intentar reactivarlos en los departamentos característicos de clientes con niños.

Una vez se consigan los resultados, se procederá a la imposición de la comunicación diferenciada en *with* y *without children* para estudiar si se consigue mejorar la reacción de los distintos hogares por los impactos más personalizados, y por tanto, el *ROI* de las campañas aumenta. Si es así, se habrá conseguido dar un paso más en la personalización de los impactos y acercar un poco más al cliente a la marca sueca.

7. Bibliografía

- Harper, M. (1961): *A new profession to aid management*. Journal of Marketing,
- Kotler, P, (2003); *Introducción al marketing*, p. 137.
- Neisbitt. J, (1984); *Megatrends: Ten new directions transforming our lives*. 1ªed.New York: WarnerBooks,
- Hernández, J. (2008): *Introducción a la minería de datos*. 1ª ed. Madrid: PEARSON PRENTICE HALL
- Joyanes. L (2014): *Big Data . Análisis de grandes volúmenes de datos en organizaciones*. 1º ed. México: Alfaomega
- Kotler. P, Armstrong. G, Saunders. J, Wong. V, Miquel. S, Bigne. J, Camara.D. (2000) () : *Introducción al marketing*. 2ª ed. Madrid: PRENTICE HALL
- Pérez. C, Santín. D. (2007): *Minería de datos. Técnicas y herramientas*. 1º ed. Madrid: THOMSON
- Pérez, L. (2009): *Técnicas estadísticas multivariantes con SPSS*. 1ª ed. Madrid: Garceta
- Porter. M, (1980), *Competitive Strategy*. 1ª ed. New York: The Free Press
- Santesmases. M, (1992)(1991): *Marketing: Conceptos y estrategias*. 2ª ed. Madrid: PIRAMIDE
- Valderrey. P, (2010): *Técnicas de segmentación de mercados*. 1ª ed. Madrid: StarBook
- IKEA Group. Información empresarial IKEA Group [sitio web]. [Consulta: julio 2015]. Disponible en: http://www.ikea.com/ms/es_ES/this-is-ikea/about-the-ikea-group/index.html y http://www.ikea.com/ms/es_ES/about_ikea/facts_and_figures/about_ikea_group/index.htm
- Demografía Portugal. Familias con hijos menores de 15 años. (sitio web) (consulta octubre 2015). Disponible en: <http://www.ine.pt>

Demografía Portugal. Información del tamaño de las familias portuguesas. (sitio web) (consulta octubre 2015). Disponible en:

<http://www.pordata.pt/Portugal/Fam%C3%ADlias+cl%C3%A1ssicas+por+n%C3%BAmero+de+indiv%C3%ADduos+segundo+os+Censos-786>

Demografía Portugal. Pirámide demográfica. (sitio web) (consulta octubre 2015). Disponible en: <http://populationpyramid.net/es/portugal/>

IKEA. 2012.Informe de sostenibilidad IKEA. (sitio web) (consulta abril 2015). Disponible en:

http://www.ikea.com/ms/es_ES/pdf/sustainability_report/sustainability_report_2012.pdf

IKEA. 2012.Informe anual IKEA Ibérica. (sitio web) (consulta abril 2015). Disponible en:

http://www.ikea.com/ms/es_ES/about_ikea/pdf/Informe_IKEAIberica_2012.pdf

IKEA. 2013.Informe anual IKEA Ibérica. (sitio web) (consulta: abril 2015). Disponible en: http://www.ikea.com/ms/es_ES/pdf/Informe_IKEA_Iberica_2013.pdf

IKEA. 2014.IKEA Group yearly summary 14. (sitio web) (consulta abril 2015). Disponible en: http://www.ikea.com/ms/en_US/pdf/yearly_summary/ikea-group-yearly-summary-fy14.pdf

Pentasoftware. 2015. Club de Fidelización IKEA Family. (sitio web) (consulta Agosto 2015). Disponible en ; <http://www.pentasoftware.es/cliente/ikea/club-de-fidelizacion-ikea-family>

8. Anexos

Anexo 1

PT_Customers		PT_Ventas	
Age	edad	CardNumber	numero de tarjeta de cliente
birthdate	cumpleaños	DEPT_CLS	departamento
businessowner	negocio	HH_ID	identificador de hogar
cardnumber	numero de tarjeta de cliente	LBL_NM	nombre del producto
childrencount	numero de hijos	MACRO_FAMILY	familia del producto
customerid	identificador del cliente	PROD_DESC	descripcion del producto
customerLanguage	lengua	PURCH_AMT	cantidad gastada
customers_effectivestore	tienda elegida	PURCH_QTY	cantidad comprada
EducationLevel	nivel de educacion	RETL_SKU_NO	referencia de producto
FamilyType	tipo de familia	TRX_DT	fecha de compra
gender	sexo	PT_Tipologia	
HH_ID	identificador de hogar	HH_ID	identificador de hogar
HomeAreaMaxRange	tamaño maximo de hogar	PURCH_AMT_Sum_total	cantidad total gastada
HomeAreaMinRange	tamaño minimo de hogar	TIPOLOGY	tipologia anterior
HomeType	tipo de casa	CardNumber	numero de tarjeta de cliente
IsValidEmail	email valido		
IsValidPostal	postal valido		
IsValidSMS	sms valido		
maritalstatus	situación familiar		
pais	pais de origen		
PMA	tienda asignada		
registrationdate	fecha de registro		
RoomsNumber	numero de habitaciones		

Anexo 2

Matriz de componentes rotados ^a															
	cocinas	ing and cook	tiles dormit	lutions y living	roky outdoor y te	ños y lavand	iebles dormit	ching y sof	on y window s	niños	work as y complement	family	asational lighti	eleria y libr	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DEPT_CLS_71__	,939	,016	,008	,015	,010	,040	,021	,013	,021	,004	,017	-,002	,007	-,015	,002
DEPT_CLS_73__	,936	,018	,003	,011	,006	,053	,022	,010	,019	,003	,012	-,005	,013	-,010	,003
DEPT_CLS_72__	,903	,019	,003	,018	,008	,041	,024	,010	,017	,002	,006	-,003	,008	,002	-,001
DEPT_CLS_79__	,754	,037	-,023	,013	,035	,098	,097	,012	,059	-,005	,031	,001	,025	-,001	-,019
DEPT_CLS_74__	,727	,024	,028	-,004	,023	,021	,000	,040	,029	,013	,017	-,003	-,011	-,009	-,008
DEPT_CLS_77__	,658	,036	,072	,025	-,010	-,001	,010	,030	,013	,012	-,010	,002	-,039	,354	,016
DEPT_CLS_75__	,553	,028	,054	,028	,008	-,016	-,018	,014	-,011	,019	-,001	-,009	-,009	,537	,004
DEPT_CLS_151__	,021	,699	,026	,024	,105	-,007	,069	,054	,024	,017	,045	,002	-,028	,014	-,020
DEPT_CLS_152__	,012	,628	-,044	,061	,213	-,024	,058	,028	,057	,036	,029	,000	-,006	,010	,002
DEPT_CLS_153__	,015	,626	,076	,037	-,049	-,003	,060	,055	,101	,026	,040	,006	-,008	-,003	-,043
DEPT_CLS_142__	,034	,606	,199	,026	-,020	,300	,014	,072	,094	,030	-,003	,001	,097	,023	,037
DEPT_CLS_141__	,054	,574	,237	,098	-,040	,212	,034	,078	,116	,047	-,004	,010	,061	,024	,035
DEPT_CLS_155__	,021	,521	,176	,048	,108	-,013	,024	,029	,050	,023	,040	,026	-,027	,030	-,014
DEPT_CLS_143__	,024	,458	,113	,049	,083	,417	,001	,054	,017	,045	-,021	,049	,002	,160	-,012
DEPT_CLS_154__	,012	,442	,139	,082	,319	,102	-,020	,044	,003	,039	,015	-,012	,136	,004	,018
DEPT_CLS_111__	,020	,203	,738	,116	,120	,056	,142	,029	,096	,016	,047	,015	,054	,011	-,030
DEPT_CLS_112__	,020	,186	,696	-,030	,062	,068	,187	,058	,096	,055	,045	,015	,069	,036	-,022
DEPT_CLS_113__	,024	,216	,588	,116	,103	,123	-,004	,022	,079	,028	,045	-,018	,007	-,006	,013
DEPT_CLS_123__	,017	,148	,394	,120	,372	,078	,032	,149	,162	,039	,002	,066	,078	-,004	-,021
DEPT_CLS_22__	,012	,065	,045	,617	-,054	,053	,059	,158	-,039	,004	,010	-,012	,017	,054	-,015
DEPT_CLS_12__	,012	,118	,097	,589	,097	-,066	,068	,188	,101	,061	,054	,023	-,059	,031	,025
DEPT_CLS_44__	,031	,075	,092	,563	,154	,080	,132	-,021	,121	-,024	,030	,000	-,059	-,033	-,017
DEPT_CLS_21__	,009	,056	-,089	,378	,074	,217	,174	,269	,105	,049	,228	,041	,134	,113	-,043
DEPT_CLS_162__	,018	,113	,064	,120	,602	,127	,001	-,009	,035	,008	,051	-,001	-,053	,037	,090
DEPT_CLS_163__	,008	,222	,026	,154	,576	,084	-,003	-,040	,044	,004	,010	,005	,179	-,006	,081
DEPT_CLS_171__	,007	,106	,063	-,178	,425	,031	,087	,124	,105	,040	,030	-,007	-,365	,103	,067
DEPT_CLS_164__	,013	,087	,104	,257	,403	,050	,012	-,006	,185	,056	,079	,062	,140	-,003	,002
DEPT_CLS_122__	,020	-,011	,051	-,115	,353	,044	,045	,095	,008	,085	-,033	-,002	,085	-,028	-,120
DEPT_CLS_76__	,231	,206	,042	,041	,064	,535	,033	,082	,067	,002	-,011	-,018	,128	,102	,022
DEPT_CLS_182__	-,004	,108	,099	,138	,117	,509	,164	-,034	,102	,065	,032	,003	,130	-,033	,018
DEPT_CLS_191__	-,001	-,029	,022	-,036	,063	,490	,013	,064	-,012	,031	,046	,020	-,069	,027	-,025
DEPT_CLS_181__	,000	,075	,043	,112	,263	,411	,015	-,001	,027	,095	,220	,012	,274	,010	,000
DEPT_CLS_183__	,061	,175	,257	,091	,012	,350	-,027	,044	,166	,050	,054	-,015	-,046	-,052	,095
DEPT_CLS_61__	,133	,015	-,019	-,006	,002	,295	,148	-,008	,284	,013	,024	-,030	-,162	,036	-,072
DEPT_CLS_41__	,023	,076	,220	,192	,007	-,020	,709	,170	-,005	,023	,028	,003	-,008	-,010	,068
DEPT_CLS_51__	,041	,093	,371	,052	,021	-,040	,650	,184	,029	,007	,036	,027	,004	,020	,045
DEPT_CLS_43__	,012	,086	,061	,343	,048	,096	,542	,102	,092	,090	,021	,012	,000	-,052	,027
DEPT_CLS_42__	,093	,030	-,114	-,020	,035	,233	,541	-,066	,095	-,007	,038	-,015	,016	,101	-,063
DEPT_CLS_81__	,046	,125	,041	,182	,020	,064	,103	,777	,071	,016	,035	-,002	,031	-,031	,006
DEPT_CLS_82__	,045	,132	,056	,152	,059	,075	,085	,771	,097	,013	,066	,012	,024	-,023	,010
DEPT_CLS_11__	,037	,088	,209	,254	,094	,017	,201	,372	,118	,022	,049	,051	-,011	,038	-,003
DEPT_CLS_78__	,011	,040	,047	-,119	,142	,239	-,021	,245	-,028	,048	,050	-,013	-,206	,244	-,041
DEPT_CLS_103__	,023	,137	,155	-,033	,066	,060	,023	,047	,703	,032	,042	,011	,125	,028	,058
DEPT_CLS_102__	,047	,131	,048	,141	,042	,065	,063	,114	,692	,053	,023	,005	-,011	,069	-,005
DEPT_CLS_101__	,019	,176	,198	,221	,265	,054	,081	,104	,495	,041	,101	,035	,050	,031	,054
DEPT_CLS_121__	,052	,038	,160	,226	,231	,169	,041	,033	,280	,034	,020	,001	,029	-,041	-,116
DEPT_CLS_92__	,007	,006	,008	,046	-,008	,085	,039	-,035	,000	,646	,034	,011	-,070	-,024	-,040
DEPT_CLS_94__	,011	,078	-,011	,062	,092	,012	,003	,024	,032	,618	,044	,005	,117	,003	,126
DEPT_CLS_91__	,006	,050	-,026	,029	,026	,044	,046	-,022	,034	,607	-,021	,006	-,031	-,029	-,001
DEPT_CLS_93__	,005	,008	,075	,033	-,005	,000	,012	-,007	,000	,589	,009	,005	,013	,003	-,050
DEPT_CLS_922__	-,003	,016	,040	-,103	,032	,012	-,034	,092	,039	,373	-,010	-,024	,114	,063	,027
DEPT_CLS_31__	,022	,050	,044	,094	,045	,101	,100	,078	,066	,022	,752	,004	,045	,020	,007
DEPT_CLS_33__	,015	,043	,059	,064	,012	,010	,012	,037	,051	,007	,725	,006	,010	,000	,027
DEPT_CLS_32__	,020	,025	,017	-,020	,017	,042	-,013	,007	,000	,012	,636	-,002	,010	-,001	-,027
DEPT_CLS_62__	,021	,053	,082	,095	,068	,053	,024	,038	,085	,018	,018	-,903	,017	,001	-,004
DEPT_CLS_131__	,006	,097	,130	,144	,130	,066	,045	,080	,123	,020	,032	,863	,010	,007	-,001
DEPT_CLS_921__	,004	,063	,092	-,045	,147	,015	-,004	,030	-,051	,062	,001	-,012	,613	,068	-,012
DEPT_CLS_105__	,000	,072	,043	-,041	,072	,120	,027	,032	,193	,082	,074	,001	,458	,027	,074
DEPT_CLS_104__	,120	,045	-,013	,082	,002	,073	,068	-,038	,116	-,001	,015	,011	,102	,798	,000
DEPT_CLS_165__	,008	,019	-,043	-,035	,122	,016	,043	-,024	,041	,014	,024	-,005	-,069	-,018	,667
DEPT_CLS_924__	,007	-,005	-,017	-,034	-,042	-,023	,042	,038	,024	,025	-,004	,007	,169	-,001	,604
DEPT_CLS_201__	,062	,048	-,152	-,134	,111	-,063	,184	,029	,124	-,002	,065	-,003	,225	-,042	-,339

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.
a. La rotación ha convergido en 8 iteraciones.

Anexo 3

DEPT_CLS	Range Area Name ENGLISH	DEPT_CLS	Range Area Name ENGLISH
11	Living room seating	111	Bedlinen
12	Living room tables	112	Quilts and pillows
21	Store and organise furniture	113	Bath textiles
22	Media solutions & accessories	121	Window solutions
31	Workspaces	122	Fabrics and accessories
41	Beds	123	Cushions, throws and chairpads
42	Wardrobes	131	Home furnishing rugs
43	Chests & other furniture	141	Cookware
44	Mirrors	142	Kitchen tools
51	Mattresses and accessories	143	Storing and washing
62	Bathroom organisers	144	Storing and washing
71	Kitchen cabinets	151	Dining and serving
72	Kitchen fronts	152	Glassware
73	Kitchen fixed interiors	153	Cutlery
74	Kitchen worktops	154	Set the table
75	Kitchen taps, sinks and sink accessories	155	Coffee and tea
76	Kitchen accessories	162	Green decoration
77	Kitchen appliances	163	Home decoration
78	Freestanding kitchens and kitchen furniture	164	Wall decoration
79	Knobs and handles	171	Outdoor
81	Dining tables	181	Small storage
82	Dining seating	182	Clothes and shoes organisation
91	Baby	183	Laundry, cleaning and sorting
92	Children's storage	191	Secondary Storage Range
93	Children 3-7 years	192	Transport and assembly range
94	Play	201	Other businesses
95	Children 8-12 years	921	Children's storage furniture
101	Lamps	922	Children's small storage
102	Integrated lighting range	32	drawest desk
103	Light sources and accessories	33	seat work
104	Seasonal lighting	105	batteries
924	child books	923	bike

Anexo 4

COCINAS AMT

DEPT_CLS_79_PURCH_AMT_Sum_Sum + DEPT_CLS_77_PURCH_AMT_Sum_Sum +
DEPT_CLS_75_PURCH_AMT_Sum_Sum + DEPT_CLS_74_PURCH_AMT_Sum_Sum +
DEPT_CLS_73_PURCH_AMT_Sum_Sum + DEPT_CLS_72_PURCH_AMT_Sum_Sum +
DEPT_CLS_71_PURCH_AMT_Sum_Sum

EATING AND COOKING AMT

DEPT_CLS_151_PURCH_AMT_Sum_Sum + DEPT_CLS_152_PURCH_AMT_Sum_Sum +
DEPT_CLS_153_PURCH_AMT_Sum_Sum + DEPT_CLS_141_PURCH_AMT_Sum_Sum +
DEPT_CLS_142_PURCH_AMT_Sum_Sum + DEPT_CLS_143_PURCH_AMT_Sum_Sum
+DEPT_CLS_155_PURCH_AMT_Sum_Sum + DEPT_CLS_154_PURCH_AMT_Sum_Sum

TEXTILES DE DORMITORIO AMT

DEPT_CLS_111_PURCH_AMT_Sum_Sum + DEPT_CLS_112_PURCH_AMT_Sum_Sum +
DEPT_CLS_113_PURCH_AMT_Sum_Sum + DEPT_CLS_123_PURCH_AMT_Sum_Sum

MEDIA SOLUTION AND LIVING ROOM TABLES

DEPT_CLS_44_PURCH_AMT_Sum_Sum + DEPT_CLS_12_PURCH_AMT_Sum_Sum +
DEPT_CLS_21_PURCH_AMT_Sum_Sum + DEPT_CLS_22_PURCH_AMT_Sum_Sum

DECORACION OUTDOOR Y TELAS POR METROS

DEPT_CLS_162_PURCH_AMT_Sum_Sum + DEPT_CLS_163_PURCH_AMT_Sum_Sum +
DEPT_CLS_171_PURCH_AMT_Sum_Sum + DEPT_CLS_164_PURCH_AMT_Sum_Sum +
DEPT_CLS_122_PURCH_AMT_Sum_Sum + DEPT_CLS_123_PURCH_AMT_Sum_Sum

BAÑOS Y LAVANDERIA AMT

DEPT_CLS_76_PURCH_AMT_Sum_Sum + DEPT_CLS_182_PURCH_AMT_Sum_Sum +
DEPT_CLS_191_PURCH_AMT_Sum_Sum + DEPT_CLS_181_PURCH_AMT_Sum_Sum +
DEPT_CLS_183_PURCH_AMT_Sum_Sum + DEPT_CLS_61_PURCH_AMT_Sum_Sum

MUEBLES DE DORMITORIO AMT

DEPT_CLS_41_PURCH_AMT_Sum_Sum + DEPT_CLS_42_PURCH_AMT_Sum_Sum +
DEPT_CLS_43_PURCH_AMT_Sum_Sum + DEPT_CLS_51_PURCH_AMT_Sum_Sum

DINING Y SOFAS AMT

DEPT_CLS_21_PURCH_AMT_Sum_Sum + DEPT_CLS_81_PURCH_AMT_Sum_Sum +
DEPT_CLS_82_PURCH_AMT_Sum_Sum + DEPT_CLS_78_PURCH_AMT_Sum_Sum +
DEPT_CLS_11_PURCH_AMT_Sum_Sum

ILUMINACION Y WINDOWS SOLUTION AMT

DEPT_CLS_61_PURCH_AMT_Sum_Sum + DEPT_CLS_103_PURCH_AMT_Sum_Sum +
DEPT_CLS_102_PURCH_AMT_Sum_Sum + DEPT_CLS_101_PURCH_AMT_Sum_Sum +
DEPT_CLS_121_PURCH_AMT_Sum_Sum

NIÑOS AMT

DEPT_CLS_91_PURCH_AMT_Sum_Sum + DEPT_CLS_92_PURCH_AMT_Sum_Sum +
DEPT_CLS_93_PURCH_AMT_Sum_Sum + DEPT_CLS_94_PURCH_AMT_Sum_Sum +
DEPT_CLS_922_PURCH_AMT_Sum_Sum

WORK AMT

DEPT_CLS_21_PURCH_AMT_Sum_Sum + DEPT_CLS_31_PURCH_AMT_Sum_Sum +
DEPT_CLS_32_PURCH_AMT_Sum_Sum + DEPT_CLS_33_PURCH_AMT_Sum_Sum

ALFOMBRAS O KOMP DE BAÑO AMT

DEPT_CLS_62_PURCH_AMT_Sum_Sum + DEPT_CLS_131_PURCH_AMT_Sum_Sum

FAMILY AMT

DEPT_CLS_105_PURCH_AMT_Sum_Sum + DEPT_CLS_921_PURCH_AMT_Sum_Sum

SEASONAL LIGHTING

DEPT_CLS_75_PURCH_AMT_Sum_Sum + DEPT_CLS_104_PURCH_AMT_Sum_Sum

PAPELERIA Y LIBRSO AMT

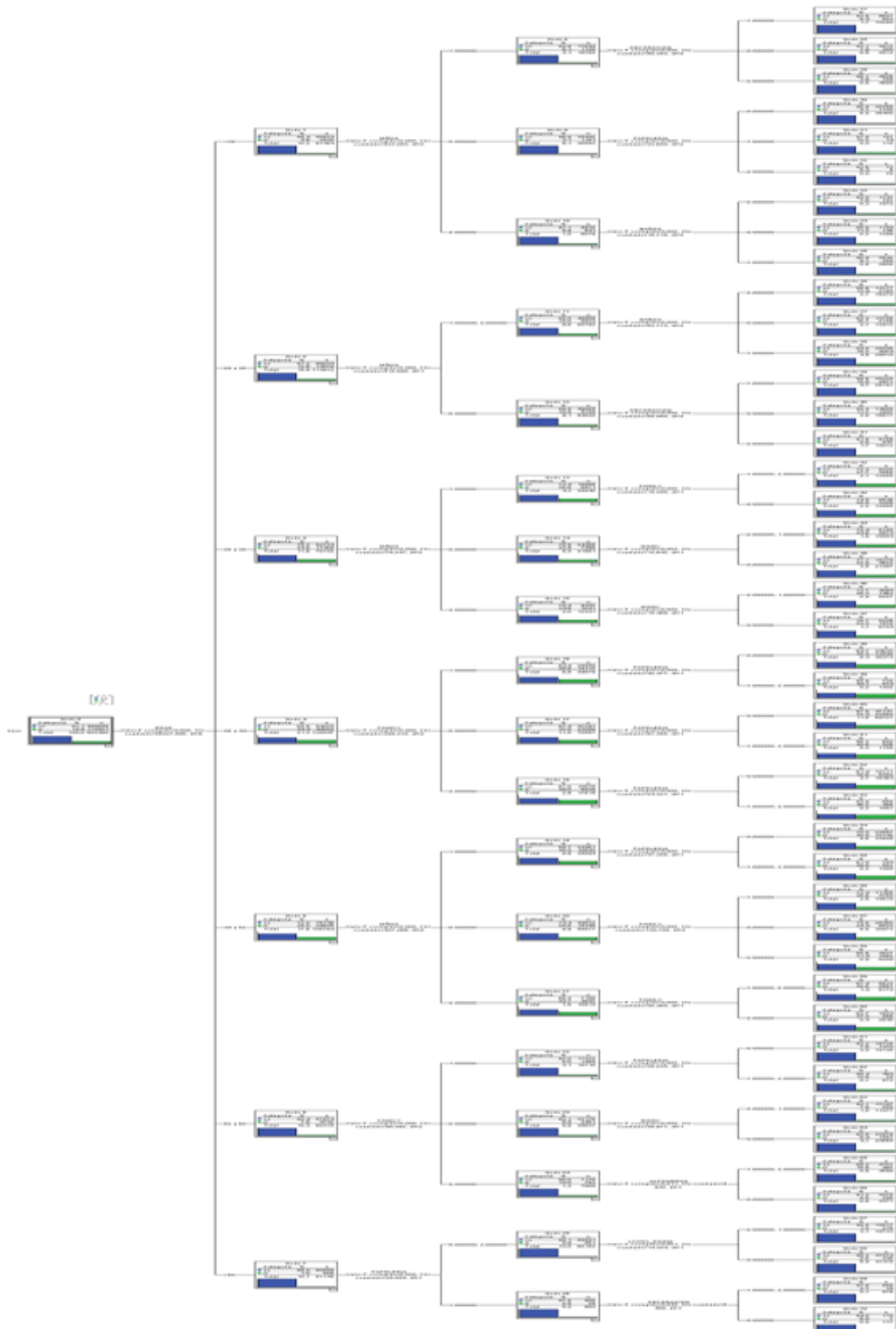
DEPT_CLS_201_PURCH_AMT_Sum_Sum + DEPT_CLS_924_PURCH_AMT_Sum_Sum

SUM AMT

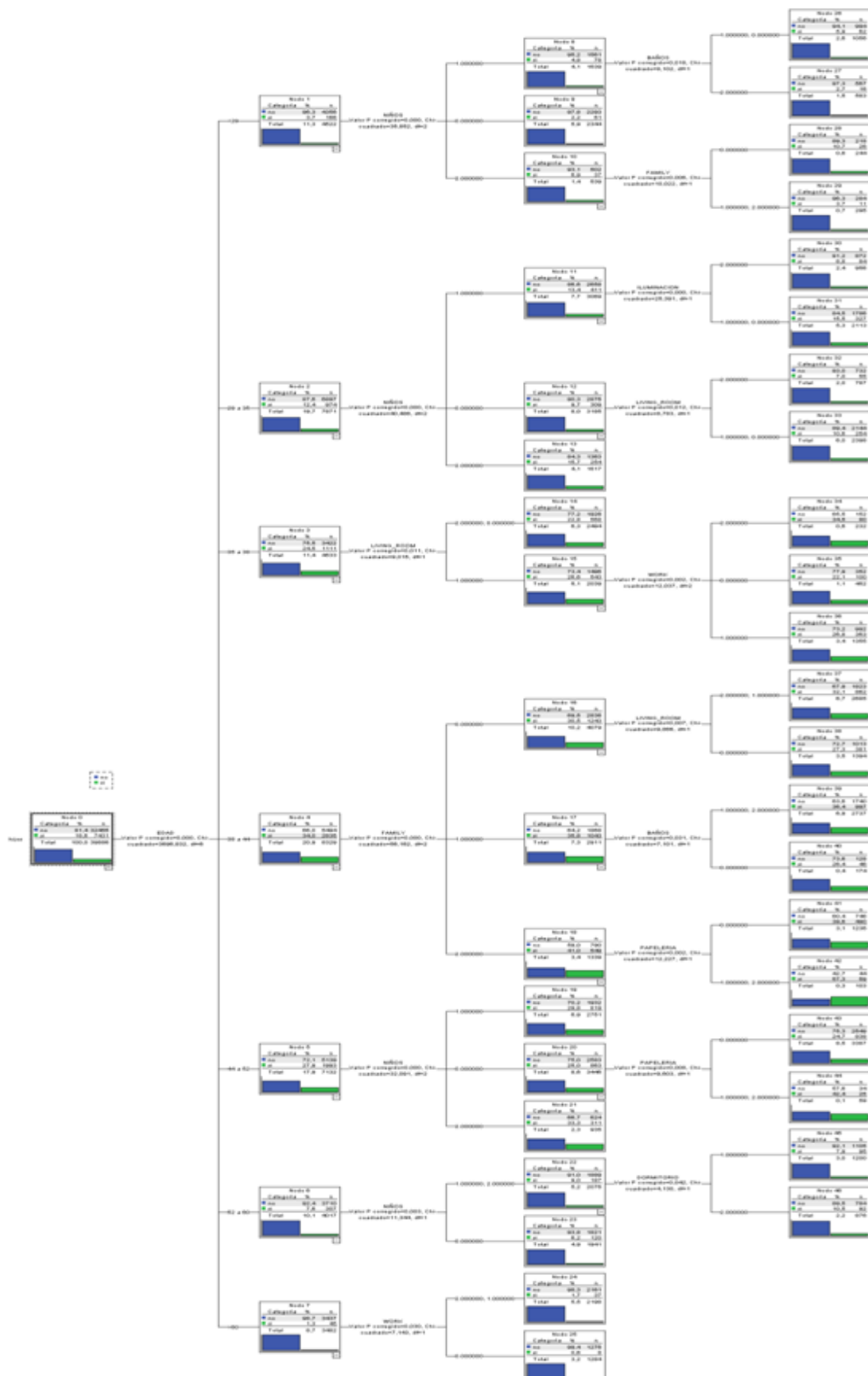
'COCINAS _amount' + 'EATING AND COOKING AMT' + 'TEXTIELS DE DORMITORIO AMOUNT' +
'MEDIA SOLUTION AND LIVING ROOM TABLES' + 'DECORACION OUTDOOR Y TELAS POR
METROS' + 'BAÑOS Y LAVANDERIA AMOUNT' + 'MUEBLES DE DORMITORIO AMT' + 'DINING
Y SOFAS AMT' + 'ILUMINACION Y WINDOWS SOLUTION AMT' + NIÑOS_AMT + WORK_AMT +
'ALFOMBRAS Y COMPLEMENTOS DE BAÑO AMT' + 'FAMILY AMT' + 'SEASONAL LIGHTINNG'
+ 'PAPELERIA Y LIBROS AMT'

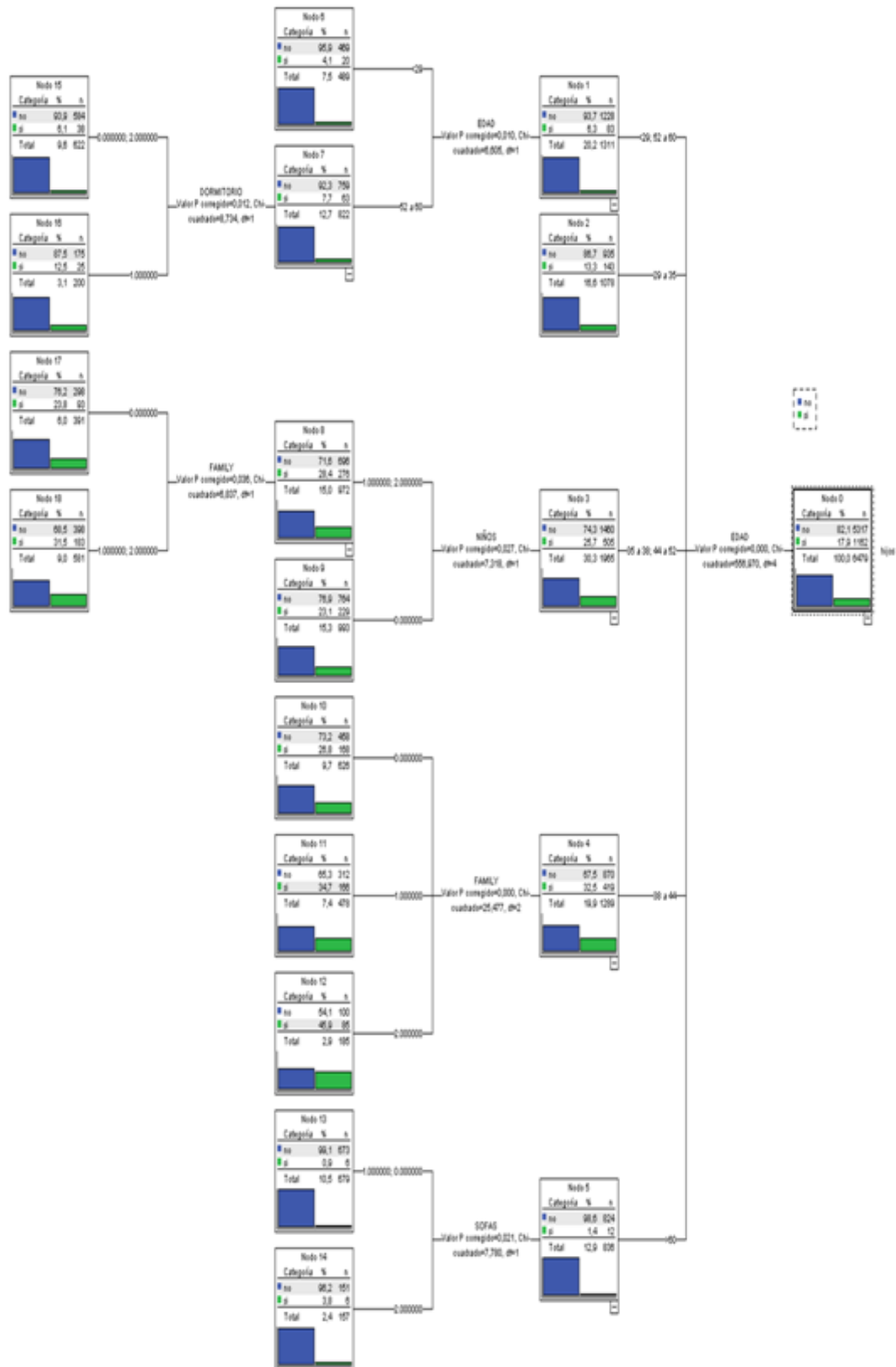
Anexo 5

Accesorios



Dormitorio





Salón

